Springer Handbook *of* Geographic Information

Kresse
Danko
Editors

2nd Edition

Springer

Springer Handbooks

Springer Handbooks provide a concise compilation of approved key information on methods of research, general principles, and functional relationships in physical and applied sciences. The world's leading experts in the fields of physics and engineering will be assigned by one or several renowned editors to write the chapters comprising each volume. The content is selected by these experts from Springer sources (books, journals, online content) and other systematic and approved recent publications of scientific and technical information.
The volumes are designed to be useful as readable desk book to give a fast and comprehensive overview and easy retrieval of essential reliable key information, including tables, graphs, and bibliographies. References to extensive sources are provided.

More information about this series at
http://www.springer.com/series/15558

Wolfgang Kresse · David Danko
Editors

Springer Handbook of Geographic Information

2nd Edition

With 620 Figures and 98 Tables

Springer

Editors
Wolfgang Kresse
University of Applied Sciences Neubrandenburg
Neubrandenburg, Germany

David Danko
Esri
Vienna, VA, USA

ISSN 2522-8692
ISSN 2522-8706 (electronic)
ISBN 978-3-030-53124-9
ISBN 978-3-030-53125-6 (eBook)
https://doi.org/10.1007/978-3-030-53125-6

Printed on acid-free paper

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

> "I do not know what I may appear to the world, but to myself I seem to have been only like a boy playing on the sea-shore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me".
> (Isaac Newton)

In this sense, the book may help us to discover a small bay in the ocean of truth and may slightly improve our insight into geographic information.

It has been almost 60 years since the phrase "geographic information system" was coined in the early 1960s. Geographers recognized the need for automation of detail-oriented processing, and computers had matured enough to handle rudimentary models of the geographic phenomenon. In the early days, geographic information systems were operated mostly in the purview of national governments and universities. As processing power and model complexity grew, and the cost of memory storage dropped, GIS slowly spread beyond governments into many disciplines and the private sector. Since the turn of the century, GIS has continued to grow rapidly, and with the advent of the Internet is now used every day even by nongeographic literate people in all walks of life.

As a typical cross-sectional science, geographic information supports many other subject areas with respect to their spatial components. The diversity of geographic information is often overlooked; one of the goals of this handbook to demonstrate this variety of applications. They range from classical subjects, such as cartography and photogrammetry, through modern fields, like Internet-based Earth browsers, to specialized systems for agriculture or health services, just to mention two of them.

Often, the term "geoinformatics" is used in place of geographic information science. This is deliberately done because the International Standardization Committee considers geographic information a specialization of information technology. Consequently, formerly prevailing paradigms have been pushed back. Cartography is no longer primarily an "engineering art". Data capture from aerial and satellite imagery is not above all precision engineering, optics, and applied mathematics. Property cadastre is not exclusively measuring art and legislation. The new philosophy is applied information technology, in particular the Internet – static and mobile, data bases, and a braiding of components from computer science. The handbook is concerned with explaining this common ground. ISO and OGC standards are referenced in many chapters as an important foundation for geographic information applications.

This second edition of the handbook follows the format of the successful first edition. Because of the rapid advancement of geographic information technology every chapter has been revised. The authors, experts in their fields, were keen to provide the latest development in the technology of their field. Several new chapters have been added: Big Data Analytics, Surveying, Building Information Modeling, Open Source GIS, and Smart Cities.

The handbook is subdivided into three parts, Basics and Computer Science, Geographic Information, and Applications. Although the structure may be quite clear, not every topic can be absolutely allocated to one of the three parts.

The first part of the book begins with a chapter about modeling; the basic concepts for abstracting real-world geographic phenomena into data for use in an information system. Other chapters in this part are linked to the basics of information technology and mathematics. The chapters about big data and the geospatial semantic web (Part 2) illuminate developments of information technology that are essential for spatial data that are often characterized by huge volumes of data, as well as multicultural and multilinguistic environments, but are still developing rapidly.

The second part addresses the specific foundations of geographic information. Chap. 6 starts with a little textbook about geographic information. Other chapters present geodesy and coordinate reference systems, data acquisition, remote sensing, metadata, geometry, and cartography, among others.

Out of the multiplicity of applications only a limited number of typical cases can be presented in the third part of the book. However, the selection focuses on the broad range of the field and stimulates the reader to gain a better understanding and, perhaps, some new ideas.

The authors, from all parts of the world, convey their distinctive perspectives of the same large field of geographic information. While in Europe geographic information is driven by legislative and organizational framework, with respect to property cadaster and planning in particular, in other parts of the world it is more technology driven.

The development of applications reaches from proprietary systems to the open-source community. The handbook allows for both. Large software vendors keep playing a predominant role in governmental systems and/or demanding developments. This is illustrated in the chapters about marine GIS and hydrography, energy suppliers, and defense. The open-source concept is addressed in the chapter of the same name. Geology, which has always been a driving force for development in cartography and geographic information, may today count for both sides, administrative systems and the open-source world.

Access to the Internet via cell-phone networks has widely abolished the distinction between static and mobile applications. Mobile applications mainly differ from static ones in their specific tasks. This topic is addressed in the chapters Geospatial Web Services, Location-Based Services, and GIS in Transportation.

Economically relevant applications are fully developed and in daily operation but only because they were preceded by research activities like those covered in several of the handbook chapters, such as in Change Detection and Marine GIS with a focus on marine ecology.

What is the distinction between a textbook and a handbook? A handbook is like a collection of many small textbooks. Every chapter conveys a good and complete summary of a subject area with references for further study. The authors have solved their tasks in different ways. Some of them have prepared the subject like a tutorial to help understand a lecture. An example is the chapter about positional accuracy improvement, including an introduction to adjustment theory in Chap. 2. Other authors explain the basics and complement these with elaborated examples, as is, e.g., done in the section about spatial databases in Chap. 3. Further on, the handbook promotes the harmonization of content and terminology, primarily in Chap. 12 about standards.

As described above, geographic information is a diversified subject that resists full documentation in one single book. We hope that our selection of topics reflects all important and many typical perspectives, and that the numerous references to other sources may help the reader to proceed where coverage by the handbook ends.

Neubrandenburg and Vienna, VA — Wolfgang Kresse
May 2022 — David M. Danko

Abbreviations

AIP	Archival Information Package
API	Application Programming Interface
ARP	aperture reference point
ASCII	American Standard Code for Information Interchange
ATS	abstract test suite
dBsm	decibels referenced to a square meter
CORBA	Common Object Request Broker Architecture
DCOM	Distributed Component Object Model
DEM	Digital Elevation Model
DIP	dissemination information package
DOP	Digital Orthophoto
DSM	Digital Surface Model
DSNU	dark signal non uniformity
DTM	Digital Terrain Model
ECEF	Earth Centered Earth Fixed
ETS	executable test suite
FITS	Flexible Image Transport System
FOM	figure of merit
FOV	Field of View
GML	Geography Markup Language
GNSS	Global Navigation Satellite System
GRP	ground reference point
HARLIE	Holographic Airborne Rotating Lidar Instrument
ICS	Implementation Conformance Statement
IERS	International Earth Rotation Service
IFOV	Instantaneous Field of View
IRI	Internationalized Resource Identifier
ISBN	International Standard Book Number
IUT	Implementation Under Test
IXIT	Implementation extra Information for Testing
JPEG	Joint Photographic Expert Group
LIDAR	Light Detection and Ranging
MCES	multichannel echo sounding
MIME	Multipurpose Internet Mail Extensions
OAIS	Open Archival Information System
OASIS	Organization for the Advancement of Structured Information
ODP	Open Distributed Processing
OGC	Open Geospatial Consortium
OSE	open systems environment
OWL	Web Ontology Language
PDI	preservation description information
PHD	phase history data

PRNU	pixel response non-uniformity
PSLR	peak side lobe ratio
RAF	Reference Architecture Foundation
RDF	Resource Description Framework
RFID	Radio Frequency Identification
SAR	Synthetic Aperture Radar
SBES	single beam echosounder
SIP	Submission Information Package
SNR	signal-to-noise ratio
SOAP	Simple Object Access Protocol
SONAR	Sound Navigation and Ranging
SUT	System Under Test
TIFF	Tagged Image File Format
TOF	time of flight
TRF	terrestrial reference frame
TRS	terrestrial reference system
UDDI	Universal Description, Discovery and Integration
UPA	ubiquitous public access
URI	Uniform Resource Identifier
VRS	Vertical Reference System
WSDL	Web Services Description Language
XML	Extensible Markup Language
XSLT	Extensible Stylesheet Language Transformations

Contents

List of Contributors

About the Editors

Wolfgang Kresse Wolfgang Kresse is an Emeritus Professor for Geographic Information Systems and Photogrammetry at the University of Applied Sciences Neubrandenburg, Germany. He also held a professorship for remote sensing and image processing at the Politechnika Koszalińska in Poland. As a visiting professor, he was affiliated to the Canada Centre for Remote Sensing in Ottawa (2002/2003) and to the University of Maine, USA (2008/2009). He is the Editor-in-chief of the German journal *PFG*, *Photogrammetry*, *Remote Sensing and Geoinformation Science*. Formerly, he worked as a software developer and customer trainer for Carl Zeiss Company with a focus on graphics, photogrammetry, and data exchange. In the ISO/TC 211 *Geographic information/Geomatics* he was the project leader of the ISO/TS 19159, parts 1 and 2, as well as the ISO 19165-1. He was also involved in the development of other ISO standards, as well as in standardization activities of the International Society for Photogrammetry and Remote Sensing (ISPRS) and of the European Spatial Data Research (EuroSDR).
Wolfgang Kresse holds a Diploma Degree in Geodesy and a Ph.D. in Digital Cartography, both of the University of Bonn, Germany. He is the primary author of the book *ISO Standards for Geographic Information*.

David Danko David Danko is a Senior Consultant for GIS standards, interoperability, and geographic metadata with Esri. He ensures that Esri software and services are interoperable and comply with the appropriate national, international, and industry standards. He has provided workshops on spatial data infrastructures, GIS standards and metadata around the globe and has participated in an advisory role in the development of many regional and national spatial data infrastructures. Before joining Esri, he worked for the U.S. Defense Mapping Agency (DMA), where he was responsible for developing and managing agency policies and programs to ensure standardization and interoperability. He also served as Project Manager for an international research and development initiative that resulted in the development of the Digital Chart of the World (DCW).

During his career at DMA he has worked in the areas of remote sensing, analytical triangulation, imagery data extraction and cartographic production.
For the last 25 years while at both Esri and DMA he served as a project leader in ISO Technical Committee 211 for the development of information technology standards. He led a team of national experts from 30 nations in the development of international metadata standard for geographic information (ISO 19115, ISO 19115-1, ISO 19115-2, ISO 19115-3, and ISO 19139).
Mr. Danko has received several awards for his work. In 2009 he was awarded the prestigious OGC Gardels Award and in 2006 he received the INCITS Merit Award for his work in standards development on the US national committee.
Mr. Danko has a B.S. (Geography/Cartography), Magna Cum Laude, from the University of Maryland.

About the Authors

Álvaro Anguix Alfaro gvSIG Association, Valencia, Spain

Andrea Antonello HydroloGIS S.r.l., Bolzano, Italy

Kristine Asch Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, Germany

Norbert Bartelme Institute of Geodesy, Graz University of Technology, Graz, Austria

Peter Baumann Jacobs University, Bremen, Germany

Matthias Becker Institute of Geodesy, Faculty of Civil and Environmental Engineering, Technical University of Darmstadt, Darmstadt, Germany

Ralf Becker Geodetic Institute and Chair for Computing in Civil Engineering and Geo Information System, RWTH Aachen University, Aachen, Germany

Ralf Bill Faculty for Agricultural and Environmental Sciences, Rostock University, Rostock, Germany

Jörg Blankenbach Geodetic Institute and Chair for Computing in Civil Engineering and Geo Information Systems, RWTH University Aachen, Aachen, Germany

Erik Borg German Remote Sensing Data Center, German Aerospace Center (DLR), Neustrelitz, Germany

Thomas Brinkhoff Institute for Applied Photogrammetry and Geoinformatics (IAPG), Jade University of Applied Sciences, Oldenburg, Germany

Jean Brodeur Recherche GéoSémantic, Sherbrooke, Québec, Canada

Robert G. Brook Hecatic Labs, Los Angeles, CA, USA; Neara, Reno, NV, USA

Aileen Buckley Esri, Redlands, CA, USA

Mario Carrera gvSIG Association, Valencia, Spain

Michael Cramer Institute for Photogrammetry (ifp), Stuttgart University, Stuttgart, Germany

David Danko Esri, Vienna, VA, USA

William F. Davenhall Geomedicine Analytics Section, Davenhall Associates, LLC, Mount Dora, FL, USA

Peter Dietrich Monitoring and Exploration Technologies, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

Kim Durante Science and Engineering Libraries, Stanford University, Stanford, CA, USA

Kian Fadaie Retired from Department of Fisheries and Oceans Canada, Ottawa, ON, Canada

Kenneth Field Esri, Redlands, CA, USA

Betsy George Spatial and Location Technologies, Oracle America, Inc., Nashua, NH, USA

Frank Gielsdorf technet GmbH Berlin, Berlin, Germany

Vladimir Golubev Faculty of Territory Development, Moscow State University of Geodesy and Cartography, Moscow, Russian Federation

Görres Grenzdörffer Faculty for Agricultural and Environmental Sciences, Rostock University, Rostock, Germany

Gerhard Gröger CPA ReDev GmbH, Siegburg, Germany

Norbert Haala Institute for Photogrammetry (ifp), University of Stuttgart, Stuttgart, Germany

Paul Hardy Cambridge, UK

John Herring Spatial and Graph, Oracle, Nashua, NH, USA

Erik Hoel Research and Development, Esri, Redlands, CA, USA

Haosheng Huang Department of Geography, Ghent University, Ghent, Belgium

Marco Hugentobler Sourcepole, Zürich, Switzerland

Knut Jetlund Department of Manufacturing and Civil Engineering, Norwegian University of Science and Technology, Gjøvik, Norway; Transport and Society, Norwegian Public Roads Administration, Hamar, Norway

Mathias Jonas International Hydrographic Organization, Monaco, Monaco

Gerhard Joos Department of Geoinformatics, Munich University of Applied Sciences, Munich, Germany

Holger Kessler Geological Modelling Systems, British Geological Survey, Nottingham, UK

Christopher Kinabrew National Network of Public Health Institutes, New Orleans, LA, USA

Jens Klump Mineral Resources, CSIRO, Perth, Australia

Angela Lausch Department of Computational Landscape Ecology, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany

Alan Leidner President of the NYC Geospatial Information Systems and Mapping Organization or GISMO, New York, NY, USA

Steve Lime MnIT Services, State of Minnesota, St. Paul, MN, USA

Roger Lott Ilkley, West Yorkshire, UK

Peter Löwe IT and eScience Department, WZB Berlin Social Science Center, Berlin, Germany

Wolfgang Kresse University of Applied Sciences Neubrandenburg, Neubrandenburg, Germany

Joan Masó CREAF Grumets research group, CREAF. Edifici C. Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain

Stephen J. Mathers Nottingham, UK

Andreas Matheus Secure Dimensions GmbH, Munich, Germany

William Meehan Utility Solutions, Esri, Redlands, CA, USA

Helena Mitasova Dept. of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC, USA

Dietmar Müller School of Geosciences, The University of Sydney, Sydney, NSW, Australia

Edward Nash DVZ Datenverarbeitungszentrum Mecklenburg-Vorpommern GmbH, Schwerin, Germany

Markus Neteler mundialis GmbH & Co. KG, Bonn, Germany

Bettina Neuhäuser Verkehrsauskunft Österreich VAO GmbH, Vienna, Austria

C. Douglas O’Brien IDON Technologies, Inc., Ottawa, Ontario, Canada

George Percivall GeoRoundtable, Crofton, MD, USA

Roland Pesch Institute for Applied Photogrammetry and Geoinformatics (IAPG), Jade University of Applied Sciences, Oldenburg, Germany

Clemens Portele interactive instruments GmbH, Bonn, Germany

Jack Reed Digital Library Systems and Services, Stanford University, Stanford, CA, USA

Charles Roswell Columbia, MD, USA

Martin Salzmann Netherland's Cadastre, Land Registry and Mapping Agency, Apeldoorn, The Netherlands

Karsten Schmidt Remote Sensing Technology Institute, German Aerospace Center (DLR), Neustrelitz, Germany

Simon Schönrock Geo- and Hydromodelling, biota – Institut für ökologische Forschung und Planung GmbH, Bützow, Germany

Markus Seifert Landesamt für Vermessung und Geoinformation, München, Germany

Jan Skaloud Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland

Christian Strobl IT Strategy, eGovernment and Applications, Bavarian State Ministry of the Environment and Consumer Protection, Munich, Germany

Jérôme Théau Université de Sherbrooke, Sherbrooke, QC, Canada

Sina C. Truckenbrodt Institute of Geography, Earth Observation, Friedrich Schiller University Jena, Jena, Germany

Paul Wessel Dept. of Earth Sciences, University of Hawaii at Mānoa, Honolulu, HI, USA

Jens Wiebensohn Premslin Farm KG, Karstädt, Germany

Jessica Wyland Marketing Communication, Esri, Redlands, CA, USA

Part A

Basics and Computer Science

From the perspective of this handbook, a special case of computer science is geographic information. Consequently, the book begins not only with mathematics, statistics, and databases, but also with data models and encoding. The fashionable term "Big Data" stands for a number of recent technologies for processing huge data volumes and is therefore a part of computer science.

The ISO 19100 series of standards for geographic information starts with data modeling. Following this concept, the handbook also starts with the modeling chapter written by John Herring, Charles Roswell, and David Danko. All three played a leading role in shaping the international standards for geographic information. The second chapter has the name mathematics and statistics and provides the basics of adjustment theory as well as geo-statistics. The background of the authors is geodesy and precision mapping (Frank Gielsdorf) and research in marine ecology (Simon Schönrock and Roland Pesch).

The database-chapter has a general part and a specific part on spatial databases written by Wolfgang Kresse and Thomas Brinkhoff respectively. The following chapter (encoding) explains the methods for the encoding of geographic feature in order to enable a transfer from one system to another. Typical approaches are GML (Geography Markup Language) and JSON (Java Script Object Notation). The author Clemens Portele was the project leader for the development of the ISO standard GML. The Big Data-chapter written by Erik Hoel addresses methods for detecting patterns, for instance distances and travel costs, in huge spatial data volumes.

1 Modeling of Geographic Information

John Herring, Charles Roswell, and David Danko

Contents

Abstract

This chapter describes methodologies for the conceptual modeling of geographic information in standards from ISO/TC 211 (Geographic information/Geomatics), OGC (Open Geospatial Consortium), or used in information management and data analytics. It begins with general concepts in data modeling, which is followed by the metaphors used in geographic informatics and processing models. Some models are database schemas, represented as object models or schemata for XML or SQL. These models are used as application schemata in object-oriented and relational databases (OODB, RDB) and in text formats for XML.

After early publications of ISO/TC 211 and OGC, with the advent of NoSQL databases and big data analytics processing on very large datasets with simpler formats, the increased use of unstructured data and simpler structures changed the direction of all information processing, including that of geographic information. More flexible data structures, with their underlying model such as JSON and YAML and dynamic programing languages, e. g., JavaScript, and CLOS became a primary component of all data management and modeling [1].

Between these two extremes are the ontology languages (e. g., OWL, RDF), which cover a wide middle ground able to describe object structure and run-time constraints. In a sense, both UML and taxonomies are limited ontology languages. Standards for these approaches have not been formalized; the first attempt was in OGC with “Features and Geometries”.

A section from the previous edition is included, which describes UML modeling (Sect. 1.3.7) which has been regularly used by geographic information standards and in other chapters.

Keywords

abstracting the real world · modeling with geometry · modeling with coverages · modeling features with taxonomies · model driven architecture · Unified Modeling Language (UML)

J. Herring (✉)
Spatial and Graph, Oracle
Nashua, NH, USA

C. Roswell
Columbia, MD, USA

D. Danko
Esri
Vienna, VA, USA
e-mail: ddanko@esri.com

1.1 Abstracting the Real World

The Earth’s geography is highly complex. Early map-makers realized that they could not represent every detail on their maps, so they focused on what they thought would provide a simplified view and address users’ needs.

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_1

When abstracting the real world for use in an information system it is helpful to make a consistent abstraction by producing a template or data model. The data model is a set of conceptual elements for representing real-world objects and processes in a digital environment. When developing models it is important to consider the type of features represented and the type of analysis to be performed.

Every profession sees the real world differently. The Department of Forestry sees roads and trails primarily as a means to get into and observe and protect the forest. The Department of Transportation has a much more in-depth focus on roads; the maintenance division is concerned with things like thickness and makeup of the base material, the type and thickness of surface materials, the width, date of last resurface, etc.; while the Traffic Management Division is concerned with the number of lanes, maximum speed limit, exit and entrance ramps, intersection control, number of vehicles per day, etc. As with the early map-makers these different views should be considered when performing abstractions and simplifications when developing geographic information focused on specific users' needs.

When individuals, groups, or organizations develop a data model, they proceed through three levels of abstraction. First, they develop a *conceptual model* limited to their domain of interest, or universe of discourse, selecting objects, relationships, behaviors, processes, and details of the real world in which they are interested. This ensures the data instantiated in the model will not only provide them all the information they need but limit data collection to just what is needed. Second, they produce a *logical model* that is an implementation-oriented (but implementation independent) construct of the objects, behaviors, processes, and details usually provided as a set of diagrams or lists. Finally, a *physical model* is produced, which defines how the model will be represented physically in a computer, as table files, databases, etc.

There are two notional views of the real world: as discrete objects or as continuous fields where property values vary in space and time. These views can be represented in any of the data model types identified below, but typically discrete objects are represented in the vector and object models, and the continuous fields are primarily represented by coverage models. Models can be static or dynamic. In a typical geographic information system (GIS), once a model is produced/used and real-world entities are instantiated in a static model, if a real-world entity changes in ways that are not in the model, a new model must be created. With dynamic models, the model can change as the real-world entities change.

There are many types of models used to incorporate geographic information into a GIS. The type of model used depends on the entities in the real world being modeled and how the modeled features will be analyzed. Some common geographic model types are listed here.

1.1.1 Computer-Aided Design (CAD)

CAD is a simple modeling system where real-world entities are represented symbolically as simple point, line, and polygon vectors focused on a graphic representation. Typically, CAD uses a local drawing coordinate system.

1.1.2 Geometry Centric Models

Coverage Model

A real-world entity is modeled as a feature that acts as a function to return values from its range for any direct position within its spatial, temporal, or spatiotemporal domain. Coverage models include raster and gridded data, digital imagery, digital elevation model (DEM), etc.

Triangulated Irregular Network (TIN)

TIN is a tessellated surface model composed of triangles typically used to model the Earth's terrain. This type of model has advantages for surface analysis, maintaining original input points enabling accuracy checks and efficient storage of surface representations. It is used in planning for Earth moving volume calculations, drainage studies, view shed analysis, and, commonly, to display shaded elevation scenes and draped with imagery for photo-realistic viewing of geographic information.

Simple Features

Real-world entities are modeled as features restricted to two-dimensional (2-D) geometry with linear interpolation between vertices having both spatial and nonspatial attributes. This is arguably the most commonly used vector-based model for geographic information today.

Topological Features

Real-world entities are modeled with similar simple features along with topological relationships.

Network Features

Real-world entities are modeled as points/nodes and lines with network topological relationships and restrictions used to model the flow of natural phenomenon, goods, and services. This type of model is primarily used in transportation, hydrology, and utility applications.

1.1.3 Object Models

Real-world entities are modeled as collections of geographic objects with complex properties, relationships, and behaviors. Other than coverage models, this type of model is the most used in robust geographic analysis.

1.1.4 Ontological Models

Real-world entities are modeled as collections of taxonomies along with their relationships. This type of model has been developing as technologies for the semantic web have developed.

1.2 Definitions

The definitions listed here have been rephrased for simplicity but, if taken strictly, they are consistent with their more complex "formal" counterparts. The semantic areas involved are listed in « » (guillemets) in the definitions, which are the stereotype delimiter for stereotypes in UML. These domains are listed from the most independent to those most dependent on the domains mentioned [2].

- Linguistics – study of the meaning, structure, expression, and use of language
- Mathematics – study of the logic and categorical structure, order, and relations that involve numbers and measurements of all conceptual types and the theory of proof
- Logic – the mathematical science of reasoning and consistency of proof
- Geometry – the mathematical study of the shape, measure, and relationships of real and conceptual things embedded in any form of conceptual reality
- Topology – the mathematical science of geometric properties that are not dependent on precise measurements and are preserved under continuous functions
- Geodesy – applied mathematical science concerned with the measurement, shape, and the most accurate possible measurement of locations and geometry associated to the Earth and the things upon it
- Informatics – applied science of the logical manipulation of information, its management, collection, query, and use
- Geomatics – informatics applied to geographic information.
- *Association:* «informatics» implementation of a relationship between digital entities
- *Attribute:* «informatics» description of some aspect of an entity
- *Barycentric coordinate system:* «geometry» coordinate system for the interior of a triangle where each point $\boldsymbol{p}$ interior to the triangle is expressed as a vector that is a combination of the three corners $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, $\boldsymbol{c}_3$ using three nonnegative numbers (a_1, a_2, a_3)

$$\boldsymbol{p} = \sum_{i=1}^{3} a_i \boldsymbol{c}_i, \quad \text{where } 0 \le a_i \le 1.0, \sum_{i=1}^{3} a_i = 1 .$$

  Note: barycentric coordinates in n-dimensions can be used for any polyhedron defined as the convex hull of $n+1$ points in general position.
- *Cellular complex:* «topology» topological complex where each cell is topologically isomorphic to a unit disc in a Cartesian coordinate space of the same dimension.
- *Class declaration:* «informatics» description of a data structure and operations that can be used to create, modify, query an object that is an instance of this class.
  Note: in a static system, the class declaration is required, a taxon that describes each object created refers to this declaration. In a dynamic system, each object is its own declaration and can be replicated by cloning (a copy of the structure that can be further modified according to the reality).
- *Clone, copy:* «informatics» (v) creation of a new object by copying the structure and behavior of an existing object or template; (n) target object of such an act of creation
  Note: once created the clone shares its "class" with its "ancestor", but, in a dynamic system, they can both "migrate" in data structure and behavior over time. In many programming languages, the cloning process is called a "copy constructor". The copy differs from the original only by its identity until some later action is completed.
- *Code lists:* «informatics» dynamic datatype consisting of a finite set of values that can be extended or shortened at any time as needed by its use.
  Note: this is like an enumeration or index that can change over time or circumstance. A taxon list for a taxonomy can be considered as a code list with an external list of definitions for all terms.
  Example: ISO produces a list of country, areas, and city names, which would have to be undated over time. The best datatype for this list would be a code list.
- *Controlled vocabulary:* «ontology» standardized set of words and phrases with definitions used to create, index, and retrieve data.
- *Coordinate reference systems:* «mathematics, geodesy» coordinate system used to present spatial information.
  Note: this includes any coordinate system used on maps (planes) or other geometric surfaces such as ellipsoids, spheroids, and geoids. A CRS is usually represented as a subset of $\mathbb{R}^n$ (ordered sets of numbers) but will not always be an $\mathbb{E}^n$, a Euclidean space.
- *Coverage:* «geomatics» function on a geographic geometry that maps each point in the geometry to a value in the range of the coverage as a function.
- *Datatype:* «informatics» entity with a set of properties (value) and operations (behavior).
  Note: the identity of a datatype is its value.
- *Dynamic object:* «informatics» identifiable digital object that can change in its value (data), relationships (associ-

ations), and behavior (methods) at any point in time as need requires.

- *Earth-centered Earth-fixed* (*ECEF*): «mathematics, geodesy» Cartesian coordinate system with its origin at the center of the Earth, the positive z-axis through the north pole, the positive x-axis through intersection of the prime meridian and the equator, and the y-axis making (x, y, z) a right-handed coordinate system.
  Note: the positive z-axis passes through the equator at 90° east longitude. Thus, positive longitude values are east and negatives are west. Positive latitudes are North, negative South.
- *Entity:* «logic, informatics» real-world thing or abstraction corresponding to an object in its digital representation.
- *Enumeration enum datatype:* «computer science, mathematics» datatype with a finite possible set of values that can be listed or enumerated.
  Note: the values in an enumeration are usually limited by the logic of the type. For example, Boolean type's values are usually true or false. In some cases, it could also be defined as "undecidable".
- *Feature:* «geomatics» representation of a real-world phenomenon (ISO 19101-1 [3]).
- *Function:* «mathematics» mapping or correspondence that associate each value in its domain set to a value in its range set.
  Note: the usual symbology for a function "f" from the domain "A" to the range "B" is "f: A → B" read as "the function 'f' from 'A' to 'B'".
- *Geoinformatics:* study, use and analysis of geospatial information.
- *Geodesic:* a curve that realizes the shortest length between any two of its points.
  Note: this is a line in a Cartesian space and a great circle on a sphere.
- *Geometric complex:* «mathematics» collection of geometries that is the realization of a topological complex.
- *Geometric realization:* «mathematics» geometric complex whose derived topological complex is isomorphic to a specific topological complex.
- *Graph database:* «informatics» datastore based on a graph model containing nodes representing entities connected in pairs by links representing relationships.
- *Information community:* «informatics» set of users with a common set of applications and data needs.
- *Isomorphic:* «mathematics» property of a mapping that preserve all properties specific to a particular theory, which implies that it can be reversed by another isomorphism.
  Note: a metric isomorphism (isometry) preserves distance, a topological isomorphism preserves all topological properties, etc.
- *Key-value pairs:* «informatics» data representation where each datum consists of a key describing the semantic purpose of the datum and a value describing its value.
  Note: the key identifies the type of entity, and the "value" is data that can act to construct the entity. The keys are often organized as a taxonomy to formalize and standardize meaning for a specific information community.
- *Metaphor:* «mathematics, linguistics» mapping between two types of conceptual models or categories describing one in a manner appropriate to the other.
  Remark: the counts or measurements of objects by associating the object properties to numbers is a categorical metaphor between the source and the integer or "real" numbers.
- *Method:* «informatics» operation associated to a set of objects or object classes.
- *Metric:* «mathematics» mathematical function that determines distances and areas in a coordinate reference system consistent with the reality of a particular surface.
  Note: except for very small area, none of these metrics are Euclidean, i. e., they are not the classical Cartesian system based on the Pythagorean formulae for triangles. For small area where everything being mapped is within a line of sight of an observation point (a few kilometers), Euclidean-Pythagorean approximations are within the usual error limits of the measurement apparatus.
- *n simplex:* «mathematics» convex hull of $n+1$ points in general position.
  Note: the boundary of an n simplex consists of $n+1$ simplices of dimension $n-1$.
- *n-dimensional simplicial complex:* «mathematics» collection of adjacent n-simplices.
- *Nested key-value pair:* «informatics» key-value pair data where values may contain other key-value pairs.
  Note: see JSON [4–6], and YAML [7] specifications.
- *NoSQL:*«informatics» "not only SQL" datastore that can represent data in means other than relations
  Note: the relational alternatives are usually key-value pairs or similar textual or binary formats, such as JSON, BSON, YAML, or some wide-table formats.
- *Object:* «informatics» instance of a datatype with identity, associations, and software methods that can model a type of entity.
- *Object class:* «informatics» set of objects that share value and association structures and methods.
- *Object template:* «informatics» description of a data structure and a set of operations that can be used to create an object instance of a class.
  Note: an object in a dynamic system is also a template since it can be cloned as needed. In a static object system, the template is the class declaration.
- *Property:* «informatics» quality, association, trait, or behavior that can be associated to an entity.

- $\mathbb{R}^n$*:* «mathematics» coordinate-space of n-tuples of real numbers.
  Note: $\mathbb{R}^n$ may not be infinite, i. e., may not range across all real numbers, nor will all points have only one coordinates. For example, in a spherical coordinate system using latitude and longitude are, respectively, limited by the $\pm 90°$ and $\pm 180°$ and the line with longitude 180° is both $-180°$ and $+180°$.
- *Simplicial complex:* «mathematics, geometry, topology» topological complex where each cell (called a "simplex") is topologically isomorphic to the convex hull of a maximal set of points in general position in a Cartesian coordinate space of the same dimension.
  Note: the first few dimensions for a simplex are isomorphic to a point, a line, a triangle, and a tetrahedron.
- *Static object:* «informatics» object that cannot change its data structure or behavior over its lifetime.
- *Taxonomy:* «informatics» organized set of terms (taxa (plural), taxon (singular)) and their definitions.
- *Topological cells:* «mathematics» open disjoint entities usually represented as sets of geometries each of a dimension defined by its geometry, 0-cells are points, 1-cell are finite curves linking one 0-cell to another, possibly the same, and so on, such that each n-cell boundary is a collection of $n-1$ cells.
  Note: an n-simplex is topologically isomorphic to an n-cell.
- *Topological complex:* «mathematics» collection of topological cells of dimension 0 to a maximal dimension "n", the dimension of the complex and where each cell in the complex has its boundary in the complex represented by the union of a set of cells of 1 smaller dimension.
- *Topology:* «mathematics» study of spatial properties invariant under one-to-one, onto, and continuous transformation
  Note: the topological relations of geometries are not changed in any transformations used in geographic processes, such as the projections from the geoid to maps or globes and those between coordinate reference systems. Because topologically the sphere is not the plane, full Earth projections have singularities at some points. Map projections such as Mercator, do not cover the poles, but are otherwise topological isomorphisms between a north and south latitude.
- *Wide table format:* «informatics» any format that mimics a relational table but does not restrict the number of columns in each row.
  Note: "wide-table" formats usually mimic "key-value pair" formats.
- $\mathbb{E}^n$*:* «mathematics» $\mathbb{R}^n$ with Euclidean geometry and a Pythagorean metric.
  Note: often referred to as a Cartesian n-space.

1.3 Modeling Features

Any data system must be able to transform information about the real world into data entities described by data models and implemented by data structures, and then to modify it as required by new or additional information. The underlying assumption is that the status of the data models so modified reflects the properties of the corresponding reality (Fig. 1.1).

The fundamental concept of geographic information modeling is the "feature". A feature is defined by ISO/TC 211 and OGC as a "representation of a real-world phenomenon" (Fig. 1.2). This parallels object-oriented programming, where a digital representation stands in for an "object" in the domain of the application being implemented. In the first object-oriented language, Simula, each "digital object" was associated to a code that "mimicked" the behavior of the corresponding "real-world object". This metaphor parallels object-oriented databases. Relational database normalization will often subdivide features across several tables, depending on its complexity. In most existing standards, the model is a classical application schema, where the objects are designed for an application, where the structure and behavior is defined by a static model (that does not change over time) and some form of schema (object (UML), relational (SQL), XML, etc.). The common assumption in these applications is that the data model is a metaphor for the real objects; concepts to be studied are reflected in the digital model.

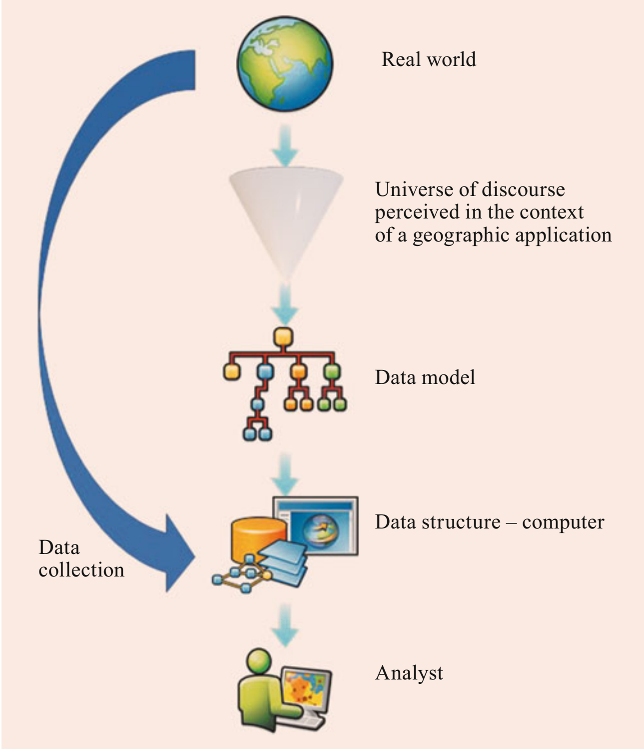


Fig. 1.1 Modeling the real world

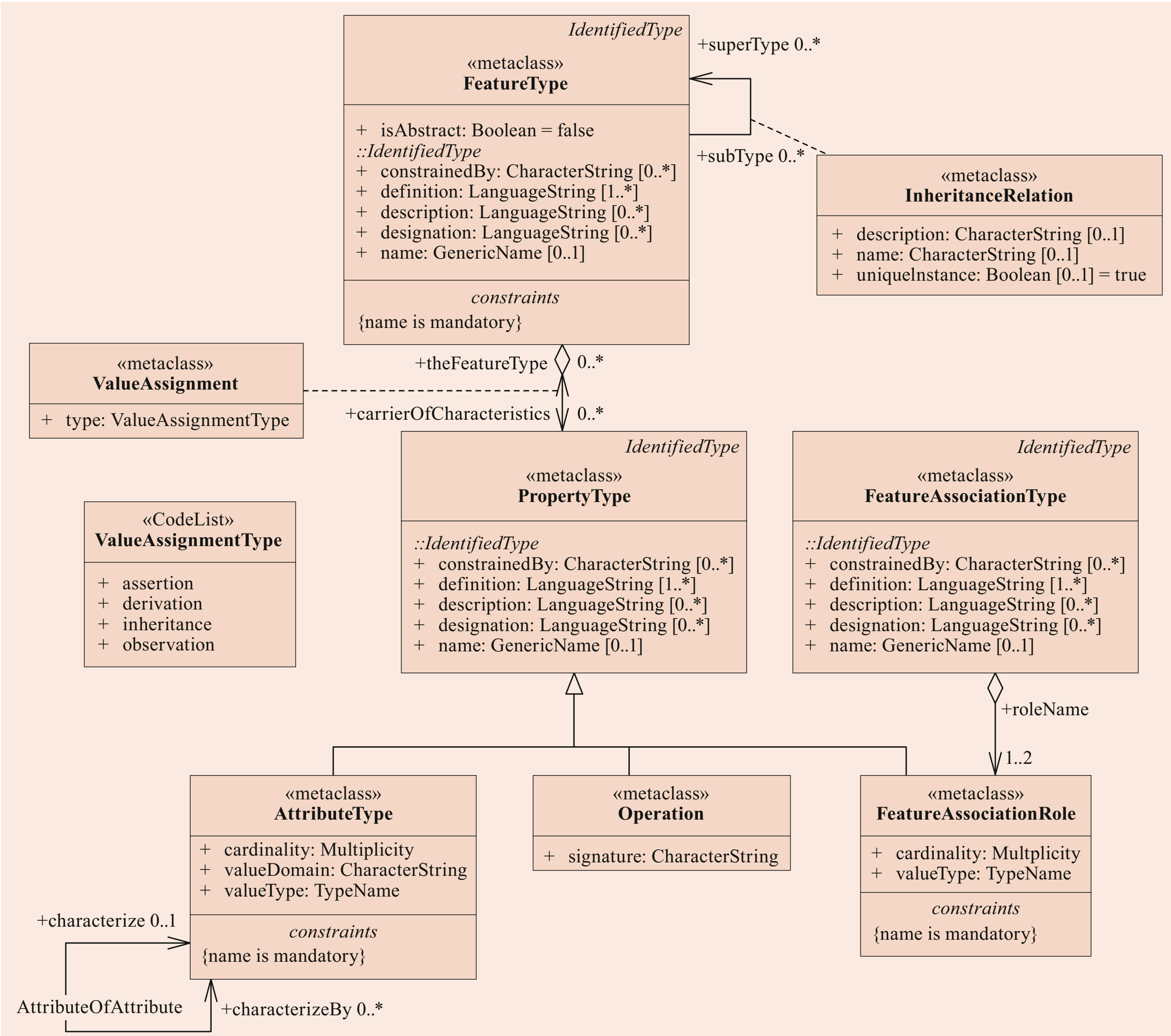


Fig. 1.2 General feature model from ISO/TC 211

The term "feature" comes from cartography, where a symbol on a map represents a real-world physical or conceptual thing.

In the various data modeling paradigms, there are two behavioral options: static or dynamic. In a static model, each object is an instance of a class that defines the structure (data) and behavior (code) of its instances. These systems are schematized. A static language or system creates object classes and modifies them during the writing of the code, e. g., in C++, before the code is compiled into machine language at compile time. In such a language, once processing starts (runtime), no new classes are created or structurally modified. Any changes in a thing in the real world would have to be reflected in the data stored as values of the digital object representing the corresponding real object. If the changes do not change structure and can be reflected by values, then the database access and edit is enough to keep up with the state of reality.

In a dynamic system, each object may begin in a specific class (possibly that of an existing object, class, or even empty class), and during its lifetime, its structure and behavior can be modified "dynamically". In a static data store, behavior can only be changed "offline", while the data and code are modified. In a dynamic data store, any object can be created or modified at any time. In a dynamic language, such as JavaScript or CLOS (common lisp object system), the class of any object can be modified during runtime, giving the object the ability to diverge from others of its class. If the structure of the data must be changed for a real-world object that has gone beyond its digital class' structure, then a dy-

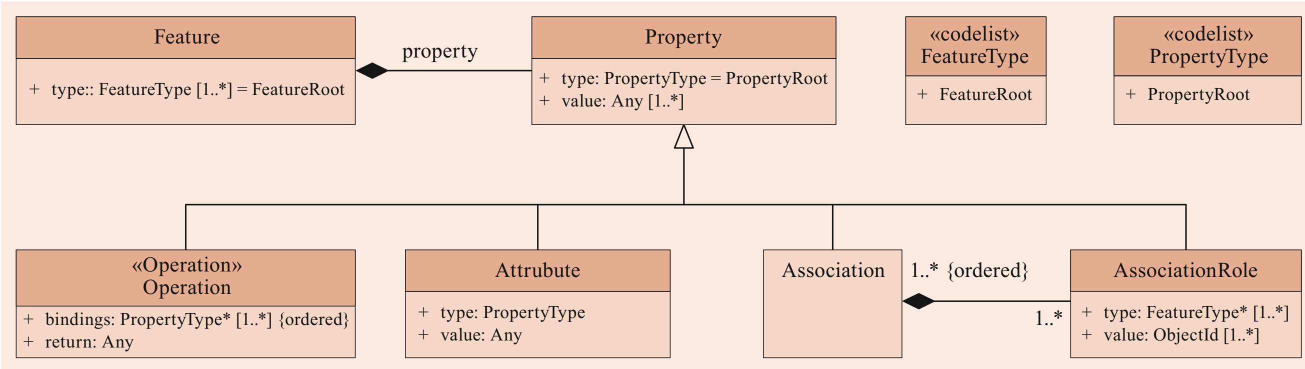


Fig. 1.3 Dynamic classes in a static language

namic digital object can be modified through a change in the object representing it.

In both static and dynamic models, the metamodel in Fig. 1.2 works. In a static model, the members of a class are defined by a static version of the figure. Each member of the class uses the same model. In a dynamic model, each object is its own class and can modify its own structure, as time can affect the structure, allowing a single object to change its structure without interfering with other objects. In a dynamic model, Fig. 1.3 is more appropriate.

The unpredictable component of the system is the ingenuity of the programmers, who can introduce policies into the programming environment that can mimic dynamic behavior in a static language or static behavior in a dynamic language, such as sharing properties via associations, as opposed to embedding them in objects and making the feature type a variable length array.

A programmer may instantiate multiple digital objects of various classes, which are then linked to one real-world object by a common "feature identity" and apply some integrity code that preserves the common properties of the "disparate object instances of the same real-world phenomenon". For example, Wheeler Dam (in northern Alabama, on the Tennessee River, constructed 1933–36), part of the TVA (Tennessee Valley Authority) has an added superstructure creating a bridge for Alabama State Route 101, like Hoover Dam (1935 on Boulder Canyon on the Nevada-Arizona border), which acts as a bridge over the Colorado River for its own access road. In a static system, the real-world object like Wheeler and Hoover dams may be described in several classes and, thus, in multiple objects.

A dynamic system (regardless of how it might be implemented) can "evolve" and allow individual "features" to avoid "pigeonholing" into a preconceived "static class". This matches the real-world better than a static class system, because no two real-world entities are identical, and can be modified or used in innovative ways at any time. So, the logic goes that each object needs to be its own class, which may require modification over its life-span.

For a single and limited application, the concept of predictable properties and behavior makes sense because the data model must match the restricted world view of the application and its concerns. These applications are usually implemented in a static system (classes, datatypes, and operations created only at compile time). In a static system, changing a class changes the properties and behaviors for all the members of the class. The system is taken offline at that time to maintain the coherence of the data.

A common static system is used in GPS-based in-car navigation systems (e. g., Garmin, TomTom, and Magellan). They all use a common graph (one-dimensional topological data structure) "node-link graph data" and some optimized variant of *Dijkstra's* graph navigation algorithm [8].

The real world is not always consistent as a single application's assumptions. In an application-independent database, applications can access data from many data sources: relational, object database, NoSQL datastores, graph databases, and various other dataset formats, such as big data formats such as key-value pairs, wide columns formats, graph, and document databases.

These data resources, often text formats (e. g., JSON for JavaScript) need defined keys that describe the semantics and datatypes of the corresponding values. In a feature set represented in key-value pair structures, the feature types and the attribute keys identify features and their properties and associations. In the 1990s, FACC (Feature and Attribute Coding Catalogue) codes used this approach. While FACC uses short keys, most data formats based on key-value pairs use words as keys that match the native language [9].

1.3.1 Model-Driven Architecture for Applications

The TC 211 and OGC approach to modeling makes use of some of the ideas concerning model-driven architecture (MDA) [10] that have been described in detail by the Object Management Group. Both organizations began their process (circa 1994) by paralleling the OMG's approach to static object modeling. The basic MDA concept is that models can be used not only to understand requirements and, thus, aid the process of system design; they can also drive the implementation, operation, and maintenance of computational systems. This concept matches the original C++ and other object-oriented languages of the 1980s and 1990s. It was also consistent with relational databases, which essentially stored datatypes in tabular formats.

Today, more flexible languages can use dynamic objects and, thus, act in many alternative data environments, as mentioned above. Moreover, more flexible data management approaches (e. g., NoSQL and BigData) tend toward more flexible data storage and ignore the "strict" data integrity of relational database stores.

1.3.2 Modeling Features with Taxonomies and Controlled Vocabularies

Taxonomies and controlled vocabularies are essentially equivalent concepts, and the taxon or term act as keys to identify the semantics of data, like an attribute or class name in an object model.

In addition, smaller vocabularies called *code lists* (sometimes "codes list", e. g., ISO 3166-1, [11]) were introduced into the ISO/TC 211 harmonized model (UML packages that are used to define object structures for the geographic information standards in the numbered series "ISO 191xx"). A code list is a small, controlled vocabulary to act as a "flexible enumeration" as values of a datatype of a class attribute. Controlled vocabularies have been used for cataloging and indexing information for many activities [12].

A key-value pair is a collection of "key" values to identify the item, followed by a delimited "(value)", which is either a datatype represented as text (as in JSON or YAML) or a text element possibly from a code list, taxonomy, or controlled vocabulary. In its simplest form, both keys and values are familiar words from a dictionary, but that can lead to confusion and cross translation issues. Words in a dictionary can have alternate meanings and multiple translations to other languages. It is simply easier to define formal lists (terms and definitions) as a taxonomy to standardize a schema structure and make it transferable to another language (and, hence, to another taxonomy in that other language). In the "best" implementations, both keys and values are derived from an appropriate collection of taxonomies.

Essentially, the code-list items in Fig. 1.3 are taxonomies for feature types and properties type (which include operations, attributes, associations, and roles). These taxonomies (whether implemented in static or dynamic languages) control the item implemented from items in Fig. 1.3.

The usual way to express this structure is to use an ABNF (Augmented Backus–Naur form) [13], ISO/IEC 14977 [14], or ABNF [15]. In general usage, the BNF use allows recursion and nesting, which allows nesting and recursion for key-value pairs, essentially the creation of a complex "object" structure. Usually, both keys and values are described in the taxonomies. For example, in representing an address the BNF may contain the following productions (using ":=" to mean is *defined to be*, "|" as *or*, parenthesis as *optional*, "< >" to denote a nonterminal, etc.):

<address> := <street address> <city>
(<state>) (<address code>) (<country>)
<street address> := <number>(<apartment>)
<street name> .

So, if <number> = "221", <apartment> = "B", <street name> = "Baker Street", <city> = "London", and <country> = "GB".

221B Baker Street, London, GB .

The resulting values are *nested key-value pairs* and are used in formats such as JSON, YAML, and WKT (well-known text) [16, 17]. Regardless of variants for BNF, each allows a parser to take values, match them to the constructs implied by the keys, and produce a full interpretation of the left-hand-side of the production, which allows the construct to be mapped in and out of any dynamic object.

Similarly, feature structures can be expressed in BNF. Standards have not yet been written on these formats. They are currently being used in NoSQL databases and in Big Data applications. It may be the case that nested key-value pairs are essentially self-explanatory, but some applications using them ask for taxonomies or some form of ontologies.

1.3.3 Modeling the Earth

The most important piece of information for a feature in geographic data is simply its location and possibly how to get there. Historically, humans have been making maps since the sixth century BC (Anaximander) and using a spherical Earth since the third century BC (Eratosthenes of Cyrene, chief librarian at the Library of Alexandria). Euclid's geometry was generalized from known Egyptian survey methods to reposition land boundaries after the annual floods of the Nile. For as long as we know, geometry and geography have been linked. Topology and graph theory are relatively new, dating from the nineteenth century with Klein, Poincaré (topology), Euler (seven bridges of Königsberg, and graph theory), and Riemann.

Beginning in the 1970s, GIS researchers were interested in creating a topological structure for all "features" captured in a two-dimensional "map space". At first, the approach was to create the topology after all the geometry for a "map" was collected. This as a general approach created a cycle of collect, structure, analyze, repeat. By about 1984, efforts began to design a system that structured the data as it was collected and merged into the existing topology.

Maps are the most common model of the geometry of the features, but applications that require highly accurate measures of the geometry on a nonplanar, nonspheroidal surface (the Earth) requires using eighteenth and nineteenth century technology and dealing with Newtonian calculus and Gaussian and Riemannian geometry. That is, differential geodesy, which deals with the shape of the Earth, which is neither flat nor spherical.

1.3.4 Modeling with Geometry

The mathematics for doing geometry on a plane or on a sphere was available by the end of the first millennium AD (1000 AD). With the advent of Newton (for calculus), Gauss, and Riemann (for differential geometry), we have all the mathematical tools we need for modern differential geodesy and doing geometric calculations on an ellipsoid, generally the WGS84 ellipsoid, which is the current reference ellipsoid for GPS location technology.

The coordinate representation of the geometry is the same as that found in engineering CAD/CAM (computer-aided design/manufacture). The difference is that engineering works in classical Euclidean geometry, sometimes denoted by $\mathbb{E}^2$ and $\mathbb{E}^3$ for the two- and three-dimensional Euclidean spaces define by Descartes (hence called Cartesian coordinates). Geographic information exists on or relatively near the Earth's surface, which is not flat. The Earth is big enough, so that, if the points on a geometry are within a few kilometers of their adjacent neighbors, the error is small. The surfaces used in geographic coordinate systems are tied to ellipsoids that approximate the "best" geodesy measurements at the time. The geometry is not so simple, and the most common approach is to embed the ellipsoidal surface in $\mathbb{E}^3$ and convert latitude, longitude (ϕ, λ), and mapped to an ("Earth-centered, Earth-fixed" $\mathbb{E}^3$ (ECEF)) X, Y, Z [18].

Some application use "spherical Mercator" which "pretends" that the Earth is close enough to a perfect sphere. GPS systems work on a WGS 84 ellipsoid, with an equatorial radius of 6,378,137 m and a polar radius of 6,356,752.3142 m, or about 42.8 km fatter than it is tall [19]. Spherical Mercator is a "little white lie", creating as much as a 15% absolute error in some measures. It is not noticeable in most graphics, but some applications might not be so loose in their error budgets. Most coordinate reference systems are defined in ISO 19111 [20, 21] and ISO 19162 [22, 23] and catalogued online at http://epsg.io/, or http://spatialreference.org/.

1.3.5 Modeling Geometry with Topology

Working directly with large sets of numerically-defined geometries can be time-consuming and so optimizing the manipulation of feature geometries is usually localized. The idea dates back the eighteenth century but was not fully developed until the late nineteenth century by Euler and Poincaré.

On a two-dimensional surface, such as a map or globe, there are three primitive geometry types: a set of points (C_0), a set of curves (C_1), and a set of areas (C_2). In a two-dimensional topological structure, these three sets of primitives are connected by a boundary operation or function "δ" that maps an area to the curves that are its limits, a curve to the points where it starts and ends (sometimes the same point), and the boundary of any point is the empty set. In mathematical (topology) theory, the surfaces have no interior holes ($C_2 \xrightarrow{\partial} C_1 \xrightarrow{\partial} C_0 \xrightarrow{\partial} \varnothing$), and the generic term is "cell". In geography, holes are allowed because lakes have islands, and the math is only slightly perturbed [24]. The optimal approach is to structure the geometry into a topological structure as it is collected and integrated in a single topological structure. The behavior of a topological editor focuses on local modification as new geometry creates new intersections. Such localized algorithms fit the newer machine model of multiple processors running in parallel.

The basic queryable spatial relations for SQL were first defined in 1994 based on the topology of objects, not on measure [25]. They were adopted in SQL in 1999 [26], ISO 19125-1, and ISO 19107. The advantage of having a topological structure from the standards gives the algorithms the advantage of navigating the topological structure instead of comparing the coordinates in complex geometries, which simply reduces the number and volume of data transfers and uses comparing topology identities instead of doing the geometric math to reproduce the original data edits. Essentially, this does the work once and then traverses a much smaller and simpler set of topological identity instead of redoing the mathematics used at collection time to create the topological structure again and again, especially as coordinate data is more accurate, as GPS takes over most of that arithmetic. Accurate positions and topology creation at collection time is more optimal than extensive human intervention. As the work to create the database is done more and more automatically and less by human interactions, the general computer-based processing to create the topology structure is done "in the cloud".

These topological relations include intersect, contains, touch, etc. The idea was put forward in the 1991 by *Egenhofer* (University of Maine, Orono) [27–30]. The advantage of using topological relations is that these calculations are invariant across coordinate transformations, which are "by definition" topological isomorphisms (all spatial relations are maintained).

The ISO standard for the geometry and topology is ISO 19107 [31]. It contains the description for a variety of types, beginning with linear interpolations as used in ISO 19125, to B-splines surfaces and solids. It also contains the implementation details for the Egenhofer spatial-query topological relations. The only similar development for spatial operators is the region connection calculus [32], which has worked out to be equivalent to the Egenhofer operators.

1.3.6 Modeling with Coverages

A coverage is a function from geographic positions on part of or all the Earth to values of some information type. The most familiar type of coverage is the "satellite" image usually related to Earth observation data. Such images are commonly binary raster, that is, a collection of pixels in a rectangular grid, each pixel containing either a gray intensity, multiple intensity based on red, green, and blue intensity, that forms a color image or by sensors that "see" in the nonvisible spectrum to take geographic measurements. Other grid interpolations can be gridded points and can then capture and using several diverse types of interpolations to represent elevation grids, or the distribution other attributes such as elevation models, temperature, wind direction, or any property that can be spatially interpolated. For discreet values, pixel (each cell has one value) structures, can be used for just about any data.

Other coverage functions are based on other geometry structures. Any randomly taken measurements at a set of points can be connected by lines by connecting each point to its nearest neighbors, usually creating a set of triangles (Fig. 1.4). Such a triangular irregular network (TIN) can use barycentric interpolation.

In Fig. 1.4, the surface elevation is represented by a set of triangles. In this case, the geometric form of the TIN mimics the elevation of the Earth. In other coverages, the mapping from the points on the Earth can represent just about anything. A forestry coverage can associate polygons to tree types to describe the forest ecosystem. Another coverage can represent species population. The coverage mechanism allows data to represent both shape and properties. For example, a census links cadaster areas and associates them to information about individuals who live there. Depending on the model chosen by the collection agency, a census could be a collection of features or a coverage. In the duality of features and coverages, the software design has options on how to express distributed information.

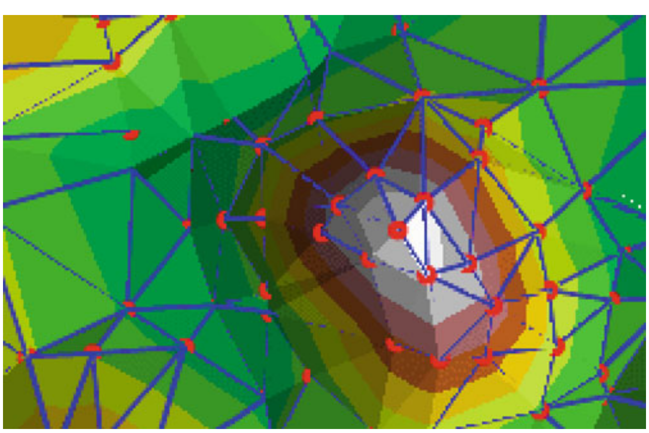


Fig. 1.4 Triangular irregular network

1.3.7 Unified Modeling Language

The Unified Modeling Language (UML) is a graphical language for specifying, constructing, and documenting systems. UML supports the construction of a variety of kinds of diagrams for modeling both the structure and behavior of systems. Conceptual modeling of information makes use primarily of structural diagrams, especially package diagrams and class diagrams, which use elements that come from the Kernel package at the core of the UML specification. A simplified description of the principal elements used in package and class diagrams is presented below. For a detailed description, see [33].

Packages

A package identifies a namespace for a group of elements contained within the package. A package is represented by a rectangle with a smaller rectangular tab at its upper left corner (Fig. 1.5). Elements contained in the package may be shown within the rectangle, in which case, the package name is placed in the tab. If contained elements are not shown, the package name is placed in the larger rectangle.

Classifiers

The fundamental element of a UML class diagram is the classifier. A classifier represents a concept within the system being modeled. It describes a set of objects that have common characteristics. Each such object is an instance of the class.

A classifier is represented by a solid rectangle containing the name of the classifier and, with the exception of the subtype *class*, a «keyword» that identifies its subtype. The

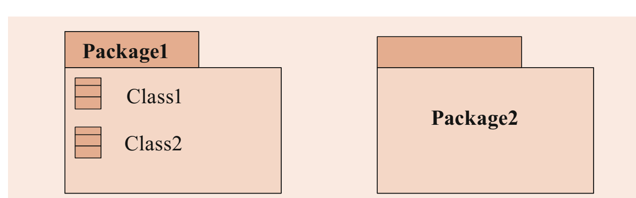


Fig. 1.5 Examples of UML packages

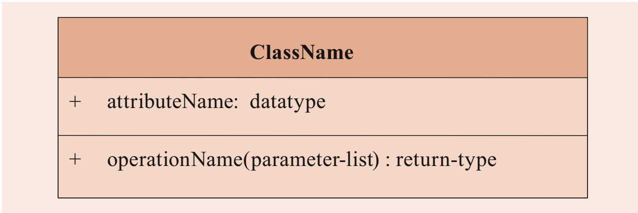


Fig. 1.6 Example of a UML class

rectangle may be divided by horizontal lines into compartments that contain features of the classifier. There are several types of classifiers.

Class

A UML class (Fig. 1.6) is a kind of classifier that has attributes and operations. A class represents a set of objects that share the same set of specifications of semantics, features, and constraints. A class may represent any set of objects, whether or not those object have a physical existence. In the case of geographic information, classes are most commonly used to represent feature types, but they may represent feature properties as well. A class is instantiated as an object, that is, as a set of attribute values and operations that describes a specific instance of the class. A class may represent lampposts or entire cities (feature types). It also may represent the ownership of parcels or the endangerment of certain species (feature properties).

The rectangle representing a class is divided into three compartments. The top compartment holds the class name and other general properties of the class; the middle compartment holds a list of attributes; the bottom compartment holds a list of operations. The attribute and operation compartments may be suppressed to simplify a diagram. Suppression does not indicate that there are no attributes or operations.

Interface

An interface describes a service offered by instances of any class that implements the interface. An interface is not instantiated in itself, nor does it specify how it is to be implemented in a class that realizes it. Rather, it describes the public behavior of a class that implements it. An interface may be implemented by several classes, and a class may implement more than one interface.

As an example, consider an interface called AreaOfVisibility. AreaOfVisibility is defined as a two-dimensional shape that outlines the area on the ground surface from which an object can be seen. The interface might be implemented by any class that represents a physical object. It could be implemented either as an operation that derives the shape from the geometric characteristics of the object and the surrounding terrain, or as an attribute that contains the shape.

An interface may be represented as a classifier diagram with the keyword «interface» in the topmost compartment

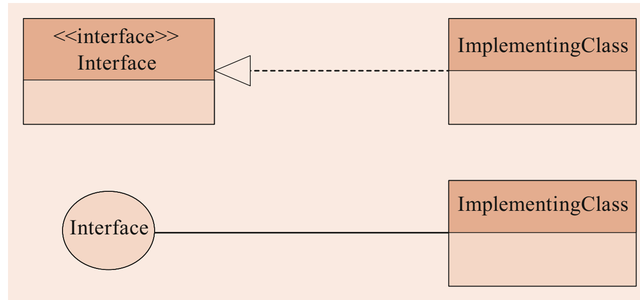


Fig. 1.7 Alternative representations of a UML interface

of the diagram. The interface diagram may be attached to the diagram representing an implementing class by a realization symbol, which is a dashed line with an open triangle at the end attached to the interface diagram (Fig. 1.7). The dependency of an implementing class upon the interface that it implements may also be shown by attaching a circle containing the interface name to the implementing class by a solid line.

Datatypes

A datatype is a kind of classifier that differs from a class only in that instances of a datatype are identified only by their value. An instance of a datatype cannot exist independently of the property whose value it provides. Datatypes include primitive predefined types and user-definable types. A datatype is identified by the keyword «dataType» in the topmost compartment of the classifier diagram.

Primitive datatypes include such things as integers and real numbers. An example of a user-defined datatype might be a combination of alphanumeric characters used to identify route numbers within national highway systems.

Enumeration

An enumeration is a kind of datatype whose instances form a list of named literal values. Both the enumeration name and its literal values are declared. An enumeration is a short list of well-understood potential values within a class. An example might be a list of the four points of the compass: east, west, north, and south.

Codelist

A codelist is a flexible enumeration specified in ISO/TS 19103 [34]. Codelists are useful for expressing a long list of potential values. An enumeration should be used when the elements of the list are completely known; a codelist should be used when only a set of likely values of the elements is known. A codelist may be extended in an application schema. The BuildingUse and LandUse codelists in Fig. 1.8 are examples. The set of codes in a codelist may be specified by a standard, such as the ISO 639-2 list of codes for identifying languages.

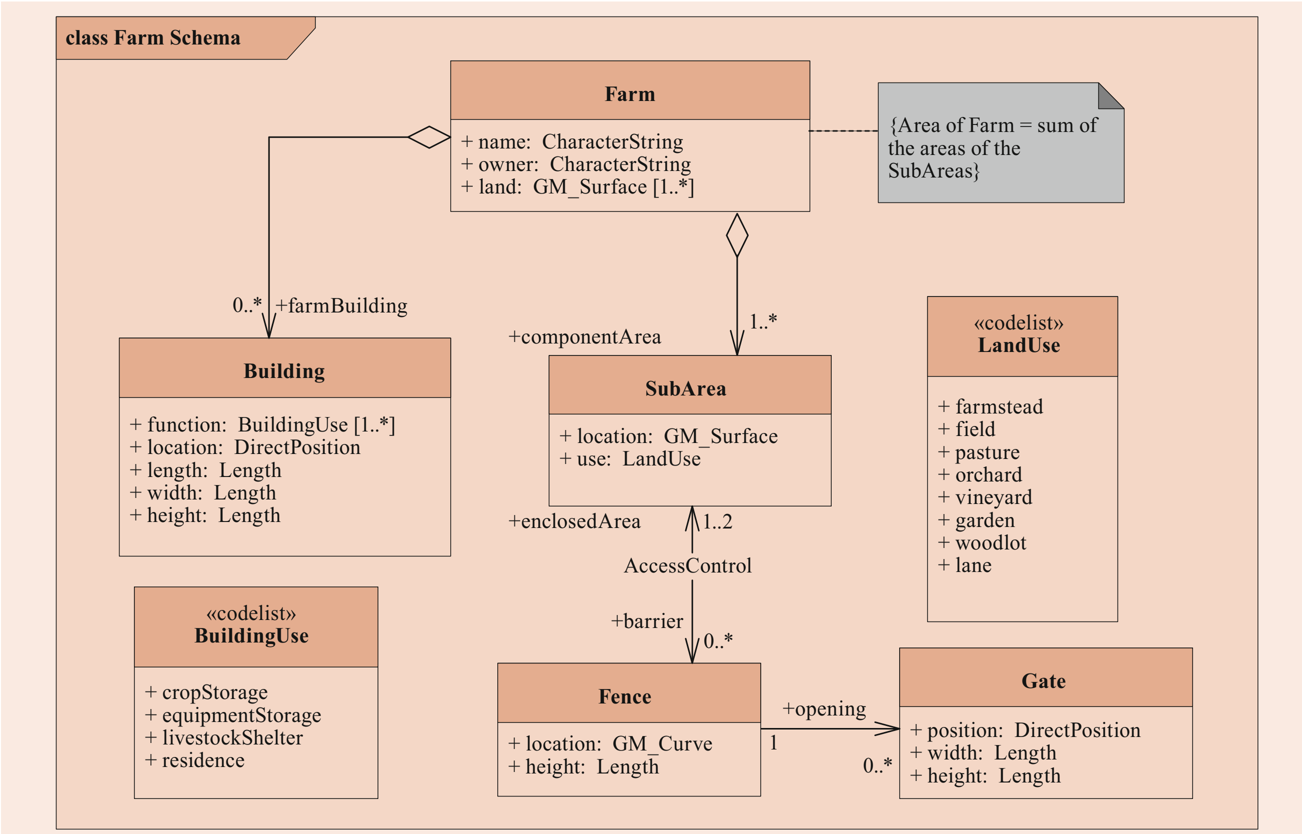


Fig. 1.8 Simple farm schema

Class Features

Attribute

An attribute represents a characteristic common to the objects of a class. Examples of attributes may be the footprint and the height of a building. An attribute is described by a string composed of elements that specify the properties of the attribute

visibility name: prop-type[multiplicity] = default,

where

- *visibility* may be public (indicated by +) or private (indicated by −). [A private element is accessible (visible) only from elements within the namespace that owns it; a public elements is visible to all elements that can access the namespace that owns it.]
- *name* is a character string that identifies the attribute. A slash (/) preceding the name indicates that the value of the attribute is derivable from the values of other model elements.
- *prop-type* identifies the datatype of the attribute.
- *multiplicity* specifies the number of values that an instance of a class may have for a given attribute. The notation for multiplicity is explained later. When multiplicity is not shown, its value is 1.
- *default* is an optional field that specifies the initial value of the attribute.

Operation

An operation represents an action that can be performed by an object. An example of an operation may be the write and read function that sets and gets the above-mentioned attributes footprint and height. An operation is described by a string composed of elements that specify the properties of the operation

visibility name (parameter-list): return-type {property-string},

where

- *visibility* may be public (indicated by +) or private (indicated by −).
- *name* is a character string that identifies the operation.
- *parameter-list* is a list of parameters, each in the form

direction parameter-name: type-expression,

where *direction* may be *in*, *out* or *inout* and defaults to *in* if the field is omitted, *parameter-name* is the name of the parameter, and *type-expression* identifies the datatype of the parameter.

- *return-type* identifies the datatype of the value returned by the operation.
- {property-string} is an optional field containing a list of property values that apply to the operation.

Association

An association (Fig. 1.9) specifies connections between instances of classes. An association is drawn as a solid line connecting the class rectangles. An association may have a name, represented as a character string placed near the line but not close to either end. Each end of an association carries information pertaining to the class at that end of the association, including its role name, its multiplicity, and its navigability. An association between two instances of the same class is represented by a line that has both ends attached to the same class rectangle.

Role Name

A role name at an association end specifies the behavior of the class at that end with respect to the class at the other end of the association. A role name is represented as a string beginning with a lower-case letter. In Fig. 1.9, the role name "owner" indicates that an instance of Person is an owner of an instance of House, while the role name "property" indicates that an instance of House is the property of an instance of Person. In data processing, the role name is used to identify the subset of instances of a class that are involved in an association. It might be used, for example, to select the owners from the set of instances of Person included in a database.

Multiplicity

Multiplicity specifies the number of instances of a class that may be associated with a class at the other end of the association. Multiplicity is expressed as a pair of integers that identify the end points of a range of values, with two periods separating the integers. An asterisk replaces the second integer if the range is unlimited. If only a single value is allowed, it may be represented by repeating that value at both ends of the range or by using a single integer in place of the range. Figure 1.9 provides the following examples.

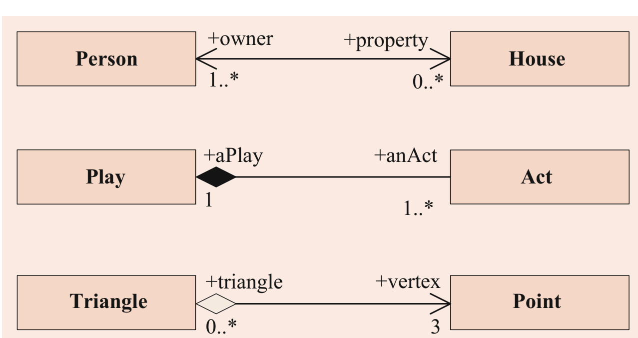


Fig. 1.9 Examples of UML associations

- An instance of House may be the property of one to many owners.
- A Person may be the owner of zero to many houses.
- A Play is composed of at least one and possibly of many instances of Act.
- An instance of Act belongs to one and only one instance of Play.
- An instance of Triangle aggregates three, and only three, points as its vertices.
- An instance of Point serves as a vertex for zero to many Triangles.

The same notation is used for multiplicity of attributes.

Navigability

Navigability describes the ability of one element to use information contained within another element. An arrowhead attached to the end of an association path indicates that navigation is allowed in the direction of the class attached at the arrowhead. In other words, information contained in that class is accessible from the class at the other end of the association. For example, Fig. 1.9 shows that it is possible to navigate from an instance of Triangle to obtain information about the instances of Point that form its vertices, but it is not possible to navigate from an instance of Point to identify the instances of Triangle for which it serves as a vertex.

Aggregation

Associations may be used to show aggregation or composition relationships between classes. An open diamond on an association end indicates that the class at that end of the association is an aggregate of instances of the class at the other end of the association. For example, an instance of the class named Triangle in Fig. 1.9 is an aggregate of three instances of the class named Point. Aggregation is considered a weak form of composition. The members of an aggregation can exist independently of the aggregation and can be members of more than one aggregation.

Composition

A closed diamond on an association end indicates that the class at that end of the association is composed of instances of the class at the other end of the association. For example, the class named Play in Fig. 1.9 is composed of one or more instances of the class named Act. Members of a composite cannot exit independently of the composite class, nor can they be members of more than one composite class.

Association Class

An association class combines the semantics of an association with those of a class. In other words, it is an association that has properties in its own right. It is represented as a class symbol attached to an association path by a dashed line

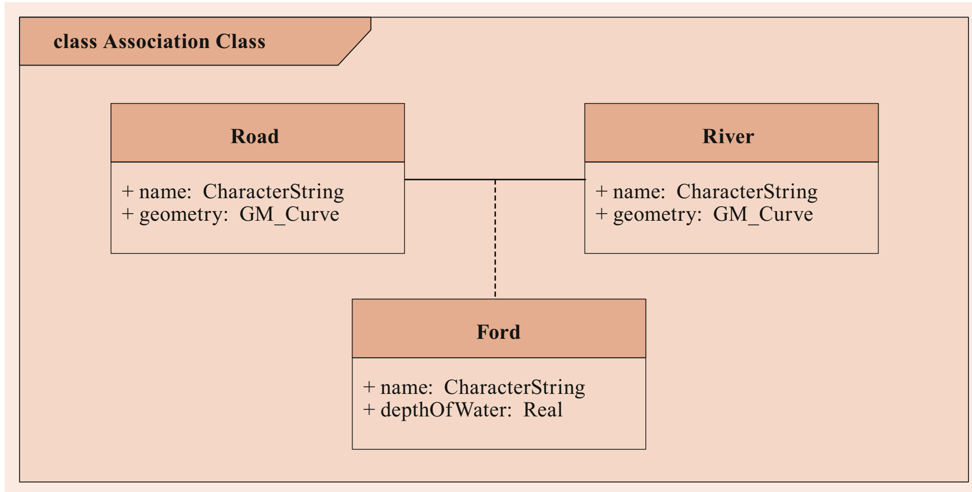


Fig. 1.10 Example of an association class

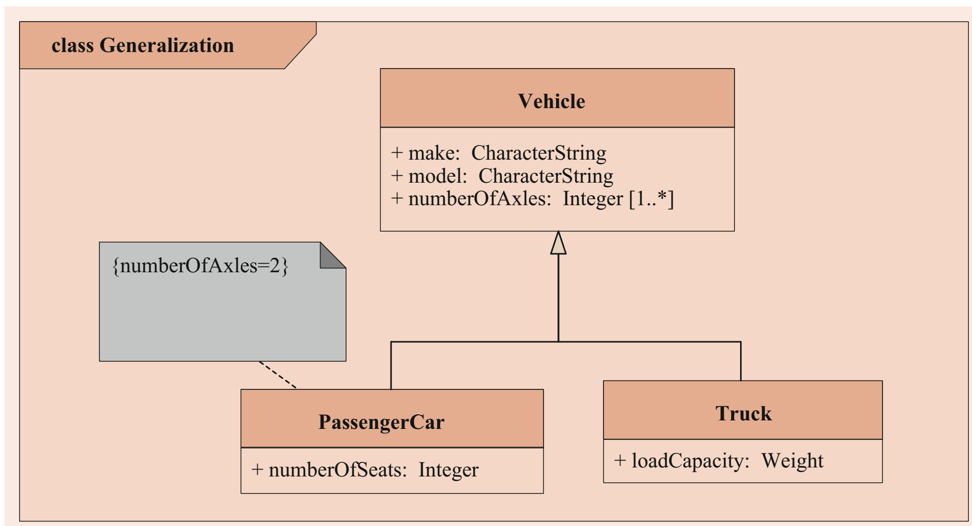


Fig. 1.11 Example of a generalization

(Fig. 1.10, in which a Ford is modeled as an association between a Road and a River).

Generalization

A generalization is a taxonomic relationship between a more general element and a more specific element. The more specific element is fully consistent with the more general element and contains additional information. An instance of the more specific element may be used where the more general element is allowed. Generalization is shown as a solid line connecting the child (the more specific element, such as a subclass) to the parent (the more general element, such as a superclass), with a large hollow triangle where the line meets the more general element. An abstract class, which has its name in italics, can only be instantiated as instances of its subclasses. Sometimes a superclass is defined as an abstract class if it has no attributes but is rather created in order to establish a clear hierarchy of the model. The resulting dataset has no objects that are named according to the abstract class.

Figure 1.11 shows an example of two generalization relationships in which each of the two subclasses of Vehicle has a third attribute in addition to the two that are inherited from Vehicle.

Individual generalization relationships may be grouped into generalization sets, each of which represents one possible way of partitioning a superclass. For example, a superclass Person might be subdivided into one generalization set on the basis of gender and into another generalization set on the basis of level of education. Each generalization set has a name that is attached to the lines connecting the subclasses in that set to the parent class.

Stereotype

Stereotypes extend the semantics of preexisting elements but do not affect their structure. A stereotype is identified by a name enclosed in guillemets («...»). An example is the stereotype «Metaclass», which identifies a classifier whose instances are themselves classes. A metaclass represents a concept at a higher level of abstraction than the classes that instantiate it.

Note

A note contains textual information. It is shown as a rectangle with the upper right corner *bent*, attached to zero or more model elements by a dashed line. Notes may be used to contain comments or constraints. The example in Fig. 1.11 contains a constraint.

Constraint

A constraint specifies a semantic condition or restriction. Although the UML specification includes an Object Constraint Language for writing constraints, a constraint may be written using any formal notation, or a natural language. A constraint is shown as a text string in braces "{ }". It is contained in a note or placed near the element to which it applies. See the example in Fig. 1.11, which constrains the value of the numberOfAxles attribute that PassengerCar inherits from Vehicle.

If the notation for an element is a text string (e. g., an attribute), the braces containing the constraint may follow the element text string. A constraint included as an element in a list applies to all subsequent elements in the list, down to the next constraint element or the end of the list.

Dependency

A dependency indicates that the implementation or functioning of one or more elements requires the presence of one or more other elements. It relates the model elements themselves and does not require a set of instances for its meaning. A dependency is shown as a dashed arrow between two model elements. The model element at the tail of the arrow (the client) depends on the model element at the arrowhead (the supplier). The kind of dependency may be indicated by a keyword in guillemets, such as «import», «refine», or «use». In the example of Fig. 1.12, Package1 has a «use» dependency upon Package2, meaning that one or more elements in Package1 use elements specified in Package2. For example, an attribute specified in Package1 might use datatypes specified in Package 2.

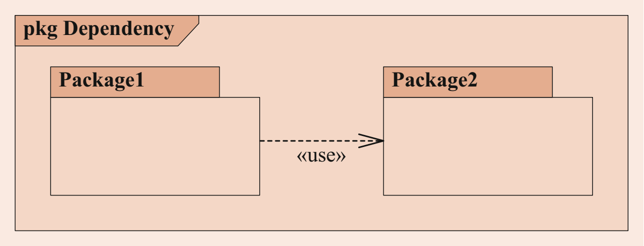


Fig. 1.12 Example of a dependency

Naming of UML Elements

Namespaces

Names of UML elements are required to be unique within the namespaces within which they reside. Thus, class names must be unique within the package within which they are specified. Attribute and operation names must be unique within the class within which they are defined. Role names must be unique within the context of the using class.

Naming Style

The UML specification provides a nonnormative set of guidelines for naming elements of UML models. They are exemplified in Figs. 1.10–1.12.

Names of classes are expressed in boldface and centered in the name compartment of the class. Each word in the class name begins with an upper-case letter, but there are no spaces between words. The name is italicized if the class is abstract.

Keywords and stereotypes are in plain typeface, placed within guillemets, and centered within the name compartment above the class name.

The first word in names of attributes, operations, parameters, and role names begins with a lower-case letter; the initial letters of subsequent words in the name begin with upper-case letters. Names of attributes and operations are left-justified within their name compartments with no spaces between words.

Conclusion

This chapter has presented some basic principles for conceptual data modeling in fields such as geographic information. It has described the use and extensions of the unified modeling language and the GFM (general feature model, ISO 19109) [35], for developing application schemas for various uses of geographic information. The current forms can alternately use taxonomies and ontologies to express more flexible dynamic models.

References

1. IETF: RFC 7946 – The GeoJSON Format. IETF, Fremont (2016). https://tools.ietf.org/html/rfc7946, Accessed 12 May 2021
2. Mark, D.M., Frank, A.U.: Cognitive and Linguistic Aspects of Geographic Space. NATO ASI Series. Kluwer Academic, Dordrecht (1991)
3. International Organization for Standardization: ISO/TC 211 Geographic information/Geomatics. ISO 19101-1:2014 – Geographic information – Reference model – Part 1: Fundamentals (2014). https://www.iso.org/standard/59164.html, Accessed 12 May 2021
4. DevDocs: JavaScript Documentation. http://devdocs.io/javascript/, Accessed 12 May 2021

5. ECMA: The JSON Data Interchange Syntax (2017). https://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf, Accessed 12 May 2021
6. Bray, T.: The JavaScript Object Notation (JSON) Data Interchange Format (2017). https://www.rfc-editor.org/info/rfc8259, Accessed 12 May 2021
7. YAML Ain't Markup Language (YAML™) Version 1.2. http://yaml.org/spec/1.2/spec.html, Accessed 12 May 2021
8. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numer. Math. **1**, 269–271 (1959). https://doi.org/10.1007/BF01386390
9. Racetin, I.: Feature definitions in feature catalogues. Cartogr. J. **52**, 67–72 (2015)
10. OMG: Model Driven Architecture (MDA). https://www.omg.org/mda/, Accessed 12 May 2021
11. International Organization for Standardization: ISO 3166-1:2013 – Codes for the Representation of Names of Countries and Their Subdivisions – Part 1: Country Codes (2013). https://www.iso.org/standard/63545.html, Accessed 12 May 2021
12. Harpring, P.: Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works, 1st edn. (Getty Research Institute, Los Angeles (2010)
13. Crocker, D., Overell, P.: Augmented BNF for Syntax Specifications: ABNF. https://tools.ietf.org/html/rfc4234, Accessed 12 May 2021
14. International Organization for Standardization: ISO/IEC 14977:1996 – Information Technology – Syntactic Metalanguage – Extended BNF. https://www.iso.org/standard/26153.html, Accessed 12 May 2021
15. Overell, P.: Augmented BNF for Syntax Specifications: ABNF (2008). https://tools.ietf.org/html/rfc5234, Accessed 12 May 2021
16. International Organization for Standardization: ISO/TC 211 Geographic information/Geomatics. ISO 19125-1:2004 – Geographic Information – Simple Feature Access – Part 1: Common Architecture. https://www.iso.org/standard/40114.html, Accessed 12 May 2021
17. OGC: 06-103r4 Simple Feature Access – Part 1: Common Architecture. http://www.opengeospatial.org/standards/sfa, Accessed 12 May 2021
18. Burkholder, E.F.: The 3-D Global Spatial Data Model: Principles and Applications, 2nd edn. CRC, Boca Raton (2017)
19. Grewal, M.S., Weill, L.R., Andrews, A.P.: Global Positioning Systems, Inertial Navigation, and Integration. John Wiley & Sons, Hoboken, New Jersey (2007)
20. International Organization for Standardization: ISO/TC 211 Geographic Information/Geomatics. ISO 19111:2019 – Geographic Information – Spatial Referencing by Coordinates. https://www.iso.org/standard/74039.html, Accessed 12 May 2021.
21. Lott, R.: OGC 08-015r2 Topic 2 – Spatial Referencing by Coordinates. http://portal.opengeospatial.org/files/?artifact_id=39049, Accessed 12 May 2021
22. International Organization for Standardization: ISO/TC 211 Geographic information/Geomatics. ISO 19162:2019 – Geographic Information – Well-known Text Representation of Coordinate Reference Systems. https://www.iso.org/standard/76496.html, Accessed 12 May 2021
23. Lott, R.: 08-015r2 The OGC Abstract Specification Topic 2 Spatial Referencing by Coordinates. http://portal.opengeospatial.org/files/?artifact_id=39049, Accessed 12 May 2021
24. Herring, J.R.: TIGRIS: A data model for an object-oriented geographic information system. In: 8th International Symposium on Computer Assisted Cartography, pp. 282–291. (1987)
25. Egenhofer, M.J., Mark, D.M., Herring, J.R.: The 9-Intersection: Formalism and Its Use For Natural-Language Spatial Predicates. Report 94-1. National Center for Geographic Information and Analysis (1994). Available at http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=68F75506EFBEDF4F516A76A051111CFC?doi=10.1.1.33.8832&rep=rep1&type=pdf, Accessed 12 May 2021
26. Egenhofer, M.J., Herring, J.R., Smith, T., Park, K.K.: A Framework for the Definition of Topological Relationships and an Algebraic Approach to Spatial Reasoning Within this Framework (91-7) (1991)
27. Egenhofer, M.J., Mark, D.M., Herring, J.R.: The 9-Intersection: Formalism and Its Use For Natural-Language Spatial Predicates. Report 94-1. National Center for Geographic Information and Analysis (1994). Available at http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=68F75506EFBEDF4F516A76A051111CFC?doi=10.1.1.33.8832&rep=rep1&type=pdf, Accessed 12 May 2021
28. International Organization for Standardization: ISO/IEC JTC 1/SC 32. Data Management and Interchange. ISO/IEC 13249-3:1999. Information Technology – Database Languages – SQL Multimedia and Application Packages – Part 3: Spatial. https://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=28054, Accessed 12 May 2021
29. Herring, J., Egenhofer, M.J., Frank, A.U.: Using category theory to model GIS applications. In: 4th International Symposium on Spatial Data Handling Zurich, Switzerland. (1990). Available at https://www.academia.edu/23673324/Using_Category_Theory_to_Model_GIS_Applications, Accessed 12 May 2021
30. Mark, D.M., Egenhofer, M.J., Herring, J.: The 9-Intersection: Formalism and Its Use for Natural-Language Spatial Predicates (1994)
31. International Organization for Standardization: ISO/TC 211 Geographic information/Geomatics. ISO 19107:2019 – Geographic information – Spatial Schema (2019). https://www.iso.org/standard/66175.html, Accessed 12 May 2021
32. Renz, J.: Qualitative Spatial Reasoning with Topological Information. Springer, Berlin (2002)
33. OMG: OMG Unified Modeling Language (OMG UML), Superstructure Version 2.4. Object Management Group, Needham (2011). available from https://www.omg.org/spec/UML/2.4/Superstructure/PDF, Accessed 12 May 2021
34. ISO/TS 19103:2015: Geographic Information: Conceptual Schema Language. ISO, Geneva (2015)
35. International Organization for Standardization: ISO/TC 211 Geographic information/Geomatics. ISO 19109:2015 – Geographic Information – Rules for Application Schema. https://www.iso.org/standard/59193.html, Accessed 12 May 2021

John Herring John R. Herring, Ph.D., has a scientific background in mathematics. Before his retirement he did research and product design for spatial database products for the Oracle Corporation. He has worked as an Adjunct Professor at the University of Maine in Orono and as an Editor for the journal GeoInformatica. Dr. Herring's involvement in standardization began in 1993 with SQL/MM. He has been engaged in ISO/TC211 and the Open Geospatial Consortium until his retirement in 2022.

Charles Roswell Charles Roswell earned a PhD in Geography in 1974. He taught at the University of Maryland from 1971 to 1980 and was employed for 27 years by the US Government as an imagery and geospatial scientist, retiring from the National Geospatial Intelligence Agency in 2008. He has more than 20 years' experience in developing national and international standards for digital geographic information.

David Danko David Danko is Senior Consultant with Esri, where he provides professional consulting services in GIS standards, interoperability, and geographic metadata. Previously, he worked for the US Defense Mapping Agency (DMA), where he was responsible for developing and managing agency policies and programs to ensure standardization and interoperability. David Danko has over 40 years experience in the fields of cartography, photogrammetry, remote sensing, and geographic information.

2 Mathematics and Statistics

Frank Gielsdorf, Simon Schönrock, and Roland Pesch

Contents

F. Gielsdorf (✉)
technet GmbH Berlin
Berlin, Germany
e-mail: frank.gielsdorf@technet-gmbh.com

S. Schönrock
Geo- and Hydromodelling, biota – Institut für ökologische Forschung und Planung GmbH
Bützow, Germany
e-mail: Simon.Schoenrock@institut-biota.de

R. Pesch
Institute for Applied Photogrammetry and Geoinformatics (IAPG), Jade University of Applied Sciences
Oldenburg, Germany
e-mail: roland.pesch@jade-hs.de

Abstract

This chapter consists of two main parts, an introduction to adjustment techniques (Sects. 2.1 and 2.2) and an overview of geostatistical methods (Sect. 2.3). The content is related to other chapters of the handbook. In particular, the adjustment technology is a foundation of many data-capture methods, and geostatistical methods are applied in marine GIS and geology.

Section 2.1 starts with an introduction to the Gauss–Markov model, discusses error propagation, and explains the role of covariance. The positional accuracy improvement, a key method for the reduction of geometrical errors present in old paper maps, is the main topic of the remainder of Sects. 2.1 and 2.2. Many related topics of positional accuracy improvement are addressed, such as datum and conformal transformation, and geometric constraints are also considered.

Section 2.3 gives a brief introduction to geostatistical analysis and modeling and introduces two approaches: universal kriging and regression kriging. Variograms for investigation and modeling are explained. Both have been applied to map benthic biotopes within the German Exclusive Economic Zone (EEZ) and coastal areas of the North Sea.

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_2

The geometrical properties of objects in geographic information systems (GIS) are almost exclusively described by point coordinates referring to a global reference frame. However, it is impossible to measure these point coordinates directly; they are the result of a calculation process. The input parameters of these calculations are measured values. Even global navigation satellite system (GNSS) receivers do not measure coordinates directly but calculate them from distance measurements to satellites.

2.1 Data Integration with Adjustment Techniques

There exist several types of measured values, for instance, distances, directions, local coordinates from maps or orthophotos, etc. Mostly, single measured values are grouped into sets of local coordinates. So, the measured values of a total station—horizontal direction, vertical direction, and distance—can be seen as a set of spherical coordinates, rows, and columns of digitized pixels as Cartesian coordinates in a local reference frame, etc. The final aim of most evaluation processes in surveying is the determination of point coordinates in a global reference frame.

However, measured values have two essential properties. Firstly, they are random variables. It is impossible to measure a value with arbitrary accuracy, which leads to the fact that any measured value contains some uncertainty. Secondly, they are redundant. Commonly, there exist more measured values than necessary to be able to calculate unique point coordinates.

A function of random variables results, again, in a random variable. Because point coordinates are functions of measurement values, they are random variables. The uncertainties contained in measured values necessarily lead to uncertainties in point coordinates.

For the unique determination of a number of coordinates, the exact same number of measured values is necessary (Fig. 2.1).

The positions of the control points A and B should be known. The distances d_1 and d_2 from the control points to the new point N have been measured. The position of the new point can be calculated by an intersection of arcs. The two unknown coordinate values of the new point, x_N and y_N, can be determined uniquely from the two measured values d_1 and d_2.

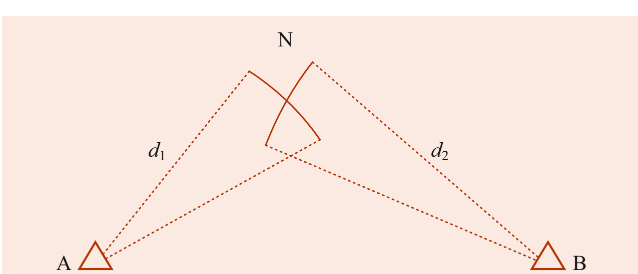


Fig. 2.1 Unique arc section

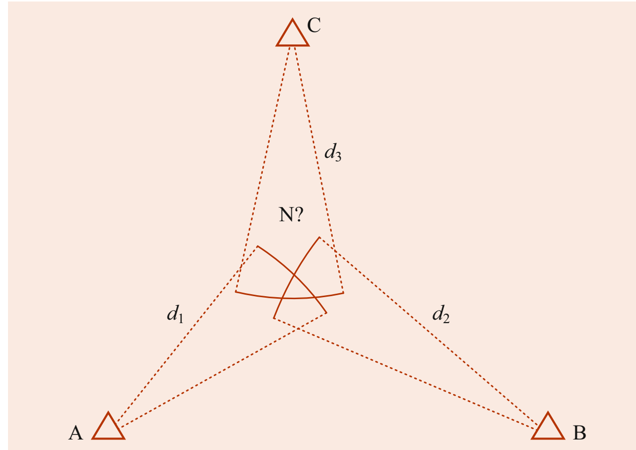


Fig. 2.2 Ambiguous arc section

However, what happens if we measure a third distance d_3 from a control point C to N (Fig. 2.2)?

A geometrical construction with a pair of compasses would yield a small triangle. The size of this triangle depends on the magnitude of the uncertainties in the measured values. For the calculation, this means that its result depends on which measured values are used. A calculation with the distances d_1 and d_2 yields a different result than a calculation with the distances d_2 and d_3.

This example raises several questions. How can one get a unique result from redundant measured values containing uncertainties? How can one quantify the accuracy of the measured values on the one hand and of the result on the other? The reply to these questions is the objective of adjustment theory.

The objectives of adjustment theory are the determination of optimal and unique output parameters (coordinates) from input parameters (measured values), which are redundant random variables, considering their accuracy, estimation of accuracy values for the output parameters, and detection and localization of blunders.

However, what is the relevance of adjustment techniques for positional accuracy improvement in GIS? Coordinates in GIS result from the evaluation of measured values. In a first step, these measured values often were local coordinates of digitized analog maps that have been transformed into a global reference frame. The global coordinates so determined describe the geometry of the GIS objects uniquely, whereby the coordinates have to be addressed as random variables. During the process of positional accuracy improvement (PAI), new measured values are introduced with higher accuracy (even global coordinates can be seen as

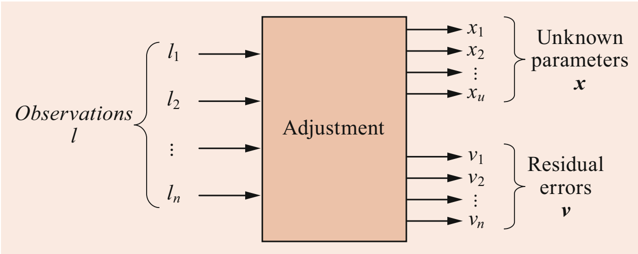


Fig. 2.3 Symbolic representation of an adjustment

special measured values). The new measured values are redundant to the already existing coordinates. Therefore, the determination of new global coordinates with improved positional accuracy is a typical adjustment problem.

However, before the application of adjustment techniques for PAI is presented in detail, it is necessary to consider the basics of adjustment theory.

2.1.1 Estimation of Parameters

Section 2.1.1 describes the process of parameter estimation on the basis of the least-squares method. Figure 2.3 is a symbolic representation of an adjustment process.

The input parameters are measured values called observations. The output parameters are the so-called unknown parameters (mostly coordinate values) and the residual errors of the observations. The number of observations is n, and the number of unknown parameters is u. Typical for an adjustment problem is the fact that the number of observations is larger than the number of unknown parameters. The difference between these two values is conterminous to the number of supernumerary observations and is called the redundancy r

$$r = n - u\,. \tag{2.1}$$

Observations are denoted by the symbol $\boldsymbol{l}$, unknown parameters by the symbol $\boldsymbol{x}$, and residual errors by the symbol $\boldsymbol{v}$. To be able to formulate adjustment problems clearly it is necessary to use vector/matrix notation. The observations, unknown parameters, and residual errors can then be grouped in vectors

$$\boldsymbol{l} = \begin{pmatrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_u \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}. \tag{2.2}$$

We only want to consider adjustment of observation equations. The basis of this model is the presentation of the observations as explicit functions of the unknown parameters. This type of approach is called a Gauss–Markov model. The observation equations in matrix notation are

$$\boldsymbol{l} + \boldsymbol{v} = \mathbf{A}\boldsymbol{x} \tag{2.3}$$

or transformed

$$\boldsymbol{v} = \mathbf{A}\boldsymbol{x} - \boldsymbol{l}\,. \tag{2.4}$$

The matrix $\mathbf{A}$ in this observation equation system is called the configuration matrix or design matrix. It describes the linear dependency between observations and unknown parameters. The observation equation system has an infinite number of solutions. To obtain the optimal solution, it is necessary to formulate an additional constraint. This additional constraint is the request that the square sum of the residual errors should be minimal. In matrix notation it is

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} \overset{!}{=} \min\,. \tag{2.5}$$

With this constraint, it is possible to formulate an extremum problem and to derive an optimal and consequently unique solution for the unknown parameters

$$\boldsymbol{x} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{l}\,. \tag{2.6}$$

The equation system (2.6) is called the normal equation system. After the normal equations have been solved, the residual errors can be calculated by inserting $\boldsymbol{x}$ into the observation equations (2.4).

2.1.2 Arithmetic Mean

Adjustment theory is very complex, and it is impossible to impart its whole content herein. For further reading please refer to [1, 2]. Therefore, we will begin with very simple special examples and try to generalize them step by step.

The simplest case of an adjustment is the calculation of an arithmetic mean value.

We consider two points A and B. The distance between these two points was measured ten times. The results of the measurement are ten observation values, which we call $l_1 \dots l_{10}$. The distance between A and B, which we call x, is the unknown parameter. Now, it would be possible to solve the problem by formulation of the equation

$$x = l_3\,,$$

in which case all the other observations would be redundant. For each observation, we could calculate a residual error v_i,

$$\begin{aligned} v_1 &= x - l_1, \\ v_2 &= x - l_2, \\ &\vdots \\ v_{10} &= x - l_{10}\,. \end{aligned} \tag{2.7}$$

The residual error v_3 would have the value 0, while the other residuals would have values different from 0. In this case, the value of x depends on the observation that we used for its determination. However, this result is not optimal. The request for an optimal result can be formulated as follows: Find exactly that solution of x for which the sum of the squares of the residual errors v_i becomes minimal.

The first step in calculating an optimal $\boldsymbol{x}$ is formulation of the observation equations

$$\boldsymbol{v} = \mathbf{A}\boldsymbol{x} - \boldsymbol{l} \ .$$

In our special case the configuration matrix $\mathbf{A}$ has a very simple structure with just one column filled with 1, and the vector $\boldsymbol{x}$ has just one row

$$\mathbf{A} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \boldsymbol{x} = (x).$$

With $\mathbf{A}$ and $\boldsymbol{l}$ we are able to calculate $\boldsymbol{x}$ as (for the derivation, see [1, 2])

$$\boldsymbol{x} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{l}.$$

We now want to apply this formula to our special problem of arithmetic mean. The matrix product $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is 10

$$\begin{array}{cc} & \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \\ (1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1) & (10). \end{array}$$

The product $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is called the normal equation matrix and is denoted by the symbol $\mathbf{N}$. We have the special case of an N-matrix with one row and one column, which is conterminous to a scalar. The inverse $(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}$ of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is equal to the reciprocal of the scalar 10

$$(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1} = \tfrac{1}{10}.$$

The expression $(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}$ (or $\mathbf{N}^{-1}$) is the cofactor matrix of the unknown parameters and is denoted by the symbol $\mathbf{Q}_{xx}$.

Cofactor matrix: $\mathbf{Q}_{xx} = \mathbf{N}^{-1}$.

The expression $\mathbf{A}^{\mathrm{T}}\boldsymbol{l}$ results in the sum of the observations

$$\begin{array}{ccc} \mathbf{A}^{\mathrm{T}}\boldsymbol{l} = & \begin{pmatrix} l_1 \\ l_2 \\ \vdots \\ l_{10} \end{pmatrix} & = \sum l_i, \\ (1 \quad 1 \quad \ldots \quad 1) & (l_1 + l_2 + \ldots + l_{10}) & \end{array}$$

so that $\boldsymbol{x}$ results in

$$\boldsymbol{x} = \tfrac{1}{10} \cdot (l_1 + l_2 + \ldots + l_{10}).$$

As we can see, the solution for the least-squares approach is the same as for the arithmetic mean.

2.1.3 Weighted Arithmetic Mean

In the arithmetic mean example, we assumed that all observations have the same accuracy. However, in reality, this assumption is mostly not applicable. In the general case, the observations have different accuracies, e. g., because of different measuring devices. For this reason, it is necessary to modify the adjustment model of the preceding section. We now have to introduce a weight for each observation. The observation weight steers the influence of each observation on the resulting unknown parameters. The greater the weight, the stronger the influence of the corresponding observation.

The weight of an observation is a function of its standard deviation. The standard deviation quantifies the accuracy of a value. The standard deviations of the observations are mostly known before the adjustment. Usually, a standard deviation is denoted by the symbol σ. The square of the standard deviation σ^2 is called the variance. The meaning of a standard deviation will be explained in the following sections. The formula for a weight p_i of an observation is

$$p_i = \frac{\sigma_0^2}{\sigma_i^2},$$

where p_i is the weight of observation i, σ_i^2 is the variance of observation i, and σ_0^2 is the variance of unit weight.

The variance of unit weight is a constant to be set. It results in weight value 1 for this variance. Often σ_0 is set to 1. If one introduces weighted observations in an adjustment, the least-squares constraint changes from

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} \stackrel{!}{=} \min$$

to

$$\boldsymbol{v}^{\mathrm{T}}\mathbf{P}\boldsymbol{v} \stackrel{!}{=} \min. \tag{2.8}$$

In expression (2.8), $\mathbf{P}$ is the weight matrix, which is a diagonal matrix of dimension $n \times n$ with the observation weights on the principle diagonal, i. e.,

$$\mathbf{P} = \begin{pmatrix} p_1 & & & 0 \\ & p_2 & & \\ & & \ddots & \\ 0 & & & p_n \end{pmatrix}.$$

The formula for the calculation of the unknown parameters $\boldsymbol{x}$ changes to

$$\boldsymbol{x} = (\mathbf{A}^\mathrm{T}\mathbf{P}\mathbf{A})^{-1}\mathbf{A}^\mathrm{T}\mathbf{P}\boldsymbol{l},$$

and the formula for the cofactor matrix of the unknown parameters $\mathbf{Q}_{xx}$ becomes

$$\mathbf{Q}_{xx} = (\mathbf{A}^\mathrm{T}\mathbf{P}\mathbf{A})^{-1}.$$

2.1.4 Adjustment with Several Unknown Parameters

The arithmetic mean represents a special case of adjustment with just one unknown parameter. In the general case, the number of unknown parameters is greater than one. This case will again be explained with a simple example.

Figure 2.4 shows points in a one-dimensional coordinate system. The coordinates x_A and x_B of the control points A and B are known. The distances between the points $d_1 \ldots d_{10}$ were measured. In demand are the coordinates of the new points $x_1 \ldots x_9$. The standard deviations are proportional to the corresponding distance

$$\sigma_i = 0.01 \cdot d_i\,. \tag{2.9}$$

In a first step we have to formulate the observation equations as

$$\begin{aligned}
d_1 + v_1 &= +1 \cdot x_1 - x_\mathrm{A} \\
d_2 + v_2 &= -1 \cdot x_1 + 1 \cdot x_2 \\
d_3 + v_3 &= -1 \cdot x_2 + 1 \cdot x_3 \\
d_4 + v_4 &= -1 \cdot x_3 + 1 \cdot x_4 \\
d_5 + v_5 &= -1 \cdot x_4 + 1 \cdot x_5 \\
d_6 + v_6 &= -1 \cdot x_5 + 1 \cdot x_6 \\
d_7 + v_7 &= -1 \cdot x_6 + 1 \cdot x_7 \\
d_8 + v_8 &= -1 \cdot x_7 + 1 \cdot x_8 \\
d_9 + v_9 &= -1 \cdot x_8 + 1 \cdot x_9 \\
d_{10} + v_{10} &= -1 \cdot x_9 + x_\mathrm{B}
\end{aligned}\,.$$

If we shift the observations to the right-hand side of the equation system, we get

$$\begin{aligned}
v_1 &= +1 \cdot x_1 - x_\mathrm{A} - d_1 \\
v_2 &= -1 \cdot x_1 + 1 \cdot x_2 - d_2 \\
v_3 &= -1 \cdot x_2 + 1 \cdot x_3 - d_3 \\
v_4 &= -1 \cdot x_3 + 1 \cdot x_4 - d_4 \\
v_5 &= -1 \cdot x_4 + 1 \cdot x_5 - d_5 \\
v_6 &= -1 \cdot x_5 + 1 \cdot x_6 - d_6 \\
v_7 &= -1 \cdot x_6 + 1 \cdot x_7 - d_7 \\
v_8 &= -1 \cdot x_7 + 1 \cdot x_8 - d_8 \\
v_9 &= -1 \cdot x_8 + 1 \cdot x_9 - d_9 \\
v_{10} &= -1 \cdot x_9 + x_\mathrm{B} - d_{10}
\end{aligned}\,.$$

From this equation system, we can directly derive the structure of our matrices

$$\mathbf{A} = \begin{pmatrix}
1 &&&&&&&& \\
-1 & 1 &&&&&&& \\
& -1 & 1 &&&&&& \\
&& -1 & 1 &&&&& \\
&&& -1 & 1 &&&& \\
&&&& -1 & 1 &&& \\
&&&&& -1 & 1 && \\
&&&&&& -1 & 1 & \\
&&&&&&& -1 & 1 \\
&&&&&&&& -1
\end{pmatrix},$$

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{pmatrix}, \quad \boldsymbol{l} = \begin{pmatrix} d_1 + x_\mathrm{A} \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \\ d_9 \\ d_{10} - x_\mathrm{B} \end{pmatrix}.$$

For the calculation of the parameter vector $\boldsymbol{x}$, the weight matrix $\mathbf{P}$ is still needed. The weights can be calculated by (2.9)

$$\mathbf{P} = \begin{pmatrix}
p_1 &&&&&&&&& \\
& p_2 &&&&&&&& \\
&& p_3 &&&&&&& \\
&&& p_4 &&&&&& \\
&&&& p_5 &&&&& \\
&&&&& p_6 &&&& \\
&&&&&& p_7 &&& \\
&&&&&&& p_8 && \\
&&&&&&&& p_9 & \\
&&&&&&&&& p_{10}
\end{pmatrix}.$$

Now, the parameter vector and the residual vector can be calculated as

$$\begin{aligned}
\boldsymbol{x} &= (\mathbf{A}^\mathrm{T}\mathbf{P}\mathbf{A})^{-1}\mathbf{A}^\mathrm{T}\mathbf{P}\mathbf{l} \\
\boldsymbol{v} &= \mathbf{A}\boldsymbol{x} - \boldsymbol{l}\,.
\end{aligned}$$

2.1.5 The Law of Error Propagation

As mentioned above, adjustment techniques have two main objectives. One task is the estimation of optimal parameter

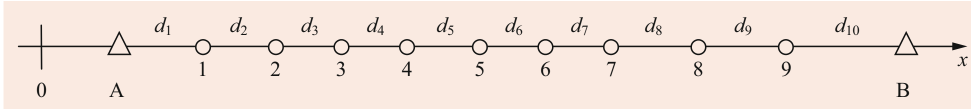


Fig. 2.4 Points in a one-dimensional coordinate reference system

values, but the second task is the estimation of parameter accuracy. The basis for this second task is the law of error propagation.

In mathematical statistics, as well as in adjustment theory, the measure for the scatter of a random value is its standard deviation. The meaning of that value shall be explained with an example.

If we measure a distance d with an observation value of 123.45 m, then this value has to be seen as a random sample of an infinite population of observations. The average value of this infinite quantity of observations is the so-called expectation value or true value. A standard deviation of $\sigma_d = \pm 2\,\text{cm}$ means that the true but unknown value of the distance d falls, with probability 67%, in the interval from 123.45 m − 2 cm to 123.45 m + 2 cm, if normally distributed observations can be assumed. The standard deviation is a measure of the average deviation of the observed value from the theoretical true value. Usually, the standard deviation is denoted by σ. The square of the standard deviation is called the variance and is denoted by σ^2.

The standard deviation is not identical to the maximal deviation from the theoretical true value. However, there exists a rule of thumb for the ratio of both values

$$\text{maximal deviation} \approx 3 \times \text{standard deviation}.$$

The input parameters of an adjustment are observations that are, in the sense of mathematical statistics, random samples of a population. The true value λ_i of an observation l_i is unknown and exists only in theory, but its standard deviation σ_i is mostly known. The output parameters of an adjustment are the parameters x_j and the residual errors v_i. Parameters as well as residual errors are functions of the observations and, therefore, also random variables.

The law of error propagation describes the propagation of accuracies for linear functions of random variables. Applying this law to an adjustment, it is possible to calculate the standard deviations of the unknown parameters and those of the residual errors.

2.1.6 Error Propagation for Linear Functions

If we have a linear function of random variables with the structure

$$x = f_1 \cdot l_1 + f_2 \cdot l_2 + \ldots + f_n \cdot l_n,$$

and the standard deviations $\sigma_1 \ldots \sigma_n$ of the random variables $l_1 \ldots l_n$ are known, then the standard deviation of the parameter x can be calculated by the formula

$$\sigma_x^2 = f_1^2 \cdot \sigma_1^2 + f_2^2 \cdot \sigma_2^2 + \ldots + f_n^2 \cdot \sigma_n^2. \tag{2.10}$$

The simplest case of a linear function is a sum. For a sum of random variables

$$x = \sum_{i=1}^{n} l_i = l_1 + l_2 + \ldots + l_n,$$

the formula for the calculation of its standard deviation is

$$\sigma_x^2 = \sum_{i=1}^{n} \sigma_i^2 = \sigma_1^2 + \sigma_2^2 + \ldots \sigma_n^2$$
$$\Rightarrow \sigma_x = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}.$$

In the case of a difference, the application of (2.10) yields

$$x = l_1 - l_2 \Rightarrow \sigma_x^2 = \sigma_1^2 + \sigma_2^2. \tag{2.11}$$

Consider Fig. 2.5 as an example. We see our one-dimensional coordinate system of the previous section. The control point A is fixed, and we want to calculate the coordinates of the new points by adding the distances

$$x_1 = x_A + d_1,$$
$$x_2 = x_A + d_1 + d_2,$$
$$\vdots$$
$$x_9 = x_A + d_1 + d_2 + \ldots + d_9.$$

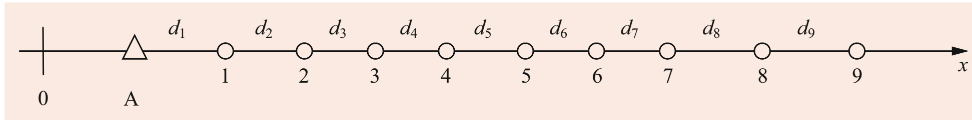
We now apply the law of error propagation to these functions and get

$$\sigma_{x1}^2 = \sigma_{d1}^2,$$
$$\sigma_{x2}^2 = \sigma_{d1}^2 + \sigma_{d2}^2,$$
$$\vdots$$
$$\sigma_{x9}^2 = \sigma_{d1}^2 + \sigma_{d2}^2 + \ldots + \sigma_{d9}^2.$$

2.1.7 The Importance of Covariances

If we applied the law of error propagation (2.11) to calculate the standard deviation of the difference $x_8 - x_7$ and compared

Fig. 2.5 Addition of distances



the result with the standard deviation of the distance d_8, then we could see that the values are not identical

$$\sigma_{x7}^2 = \sigma_{d1}^2 + \sigma_{d2}^2 + \sigma_{d3}^2 + \sigma_{d4}^2 + \sigma_{d5}^2 + \sigma_{d6}^2 + \sigma_{d7}^2,$$
$$\sigma_{x8}^2 = \sigma_{d1}^2 + \sigma_{d2}^2 + \sigma_{d3}^2 + \sigma_{d4}^2 + \sigma_{d5}^2 + \sigma_{d6}^2 + \sigma_{d7}^2 + \sigma_{d8}^2.$$

Why is this? Does this imply $\sigma_{x8-x7}^2 = \sigma_{x8}^2 + \sigma_{x7}^2$?

The answer is that we neglected the covariance between the random values x_7 and x_8. However, before we explain calculations with covariances, we want to interpret the standard deviations of the distances d_i and those of the coordinates x_i. The standard deviations of the coordinates σ_{xi} represent the absolute accuracy of the coordinates relative to the reference frame. On the other hand, the standard deviations of the distances σ_{di} represent the accuracy of the coordinates relative to each other.

However, what does covariance mean? If two calculated random values are functions of partially the same random variable arguments, they are stochastically dependent. The degree of their stochastic dependency is quantified by their covariance.

The parameters x_7 and x_8 are stochastically dependent because they are functions of partials of the same random variables. The distances d_1 to d_7 are arguments of both functions; just d_8 is an argument of only one of the two functions.

However, how can we incorporate these dependencies into the error propagation? This problem can be solved by a generalization of the law of error propagation. The general form of the law of error propagation can be represented in matrix notation. If there is a system of linear equations describing the functional dependency of parameters x_j on the arguments l_i

$$\boldsymbol{x} = \mathbf{F} \cdot \boldsymbol{l}, \qquad (2.12)$$

and the standard deviations of the arguments l_i are known, then the variances and covariances of the parameters x_j can be calculated by the formula

$$\mathbf{C}_{xx} = \mathbf{F} \cdot \mathbf{C}_{ll} \cdot \mathbf{F}^{\mathrm{T}}. \qquad (2.13)$$

In (2.13), the functional matrix $\mathbf{F}$ contains the coefficients of the linear functions. The matrix $\mathbf{C}_{ll}$ is called the covariance matrix of observations and contains the variances of observations on its principal diagonal and their covariances on its secondary diagonals, which run from the lower left to the upper right corner. In the most common case of stochastically independent observations, $\mathbf{C}_{ll}$ is a diagonal matrix; $\mathbf{C}_{xx}$ is the covariance matrix of the unknown parameters and contains their variances and covariances

$$\mathbf{C}_{ll} = \begin{pmatrix} \sigma_{l1}^2 & & & \\ & \sigma_{l2}^2 & & \\ & & \ddots & \\ & & & \sigma_{\mathrm{ln}}^2 \end{pmatrix},$$

$$\mathbf{C}_{xx} = \begin{pmatrix} \sigma_{x1}^2 & \mathrm{cov}(x_1, x_2) & \dots & \mathrm{cov}(x_1, x_u) \\ \mathrm{cov}(x_1, x_2) & \sigma_{x2}^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \mathrm{cov}(x_1, x_u) & \dots & \dots & \sigma_{\mathrm{xu}}^2 \end{pmatrix}.$$

Covariance matrices are always quadratic and symmetric.

Now, we want to apply the general law of error propagation to our example. First, we have to build the functional matrix $\mathbf{F}$. Because the functions are simple sums, $\mathbf{F}$ contains just the values 1 and zero

$$\mathbf{F} = \begin{pmatrix} 1 & & & & & & & & \\ 1 & 1 & & & & & & & \\ 1 & 1 & 1 & & & & & & \\ 1 & 1 & 1 & 1 & & & & & \\ 1 & 1 & 1 & 1 & 1 & & & & \\ 1 & 1 & 1 & 1 & 1 & 1 & & & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix},$$

$$\mathbf{C}_{ll} = \begin{pmatrix} \sigma_{d1}^2 & & & & & & & & \\ & \sigma_{d2}^2 & & & & & & & \\ & & \sigma_{d3}^2 & & & & & & \\ & & & \sigma_{d4}^2 & & & & & \\ & & & & \sigma_{d5}^2 & & & & \\ & & & & & \sigma_{d6}^2 & & & \\ & & & & & & \sigma_{d7}^2 & & \\ & & & & & & & \sigma_{d8}^2 & \\ & & & & & & & & \sigma_{d9}^2 \end{pmatrix}.$$

In the result of the calculation, we get the covariance matrix $\mathbf{C}_{xx}$, which is fully allocated (fully equipped).

Now, we can use elements of $\mathbf{C}_{xx}$ for a further calculation to get the standard deviation of the coordinate difference $x_8 - x_7$. The functional matrix $\mathbf{F}$ for that case is simple again

$$\mathbf{F} = \begin{pmatrix} -1 & 1 \end{pmatrix}.$$

The $\mathbf{C}_{ll}$ matrix contains the variances of x_7 and x_8, as well as their covariance from the previous calculation

$$\mathbf{C}_{ll} = \begin{pmatrix} \sigma_{x7}^2 & \text{cov}(x_7, x_8) \\ \text{cov}(x_7, x_8) & \sigma_{x8}^2 \end{pmatrix}.$$

If we solve the matrix equation for the general law of error propagation in a symbolic way, then we get the expression

$$\sigma_{x8-x7}^2 = \sigma_{x7}^2 + \sigma_{x8}^2 - 2 \cdot \text{cov}(x_7, x_8).$$

As we can see, this formula, which does not neglect the covariance between dependent random variables, yields the correct result.

2.1.8 Adjustment and Error Propagation

If we consider the calculation formula for the unknown parameters in an adjustment, then we can see that the parameters are linear functions of the observations, i. e.,

$$\boldsymbol{x} = \underbrace{(\mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{P}}_{\mathbf{F}} \cdot \boldsymbol{l}.$$

The expression $(\mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{P}$ in this formula is equivalent to the functional matrix $\mathbf{F}$ in (2.12). This approach allows for the application of the law of error propagation to calculate the covariance matrix of the unknown parameters.

The basis of further calculations is the empirical standard deviation of unit weight. The formula for its calculation is

$$s_0 = \sqrt{\frac{\boldsymbol{v}^{\mathrm{T}}\mathbf{P}\boldsymbol{v}}{r}}, \tag{2.14}$$

where $\boldsymbol{v}$ is the residual vector, and r is the redundancy. The redundancy is the number of surplus observations, equal to the difference between the number of observations n and the number of unknown parameters u, i. e.,

$$r = n - u. \tag{2.15}$$

The value s_0 can be interpreted as the empirical standard deviation of an observation with weight $p = 1$.

A standard deviation σ should not be mistaken for an empirical standard deviation s. A standard deviation is a constant value and an input parameter of an adjustment calculation, whereas an empirical standard deviation is a random variable, which is estimated as an output parameter of an adjustment. After an adjustment calculation, the empirical standard deviation s_0 should be approximately equal to the standard deviation σ_0, which means that the quotient s_0/σ_0 should be in the interval 0.7–1.3. If s_0 is too small, the a-priori assumption of observational accuracy (σ_i) was too pessimistic. Too large a value of s_0 often indicates that there is a blunder among the observations.

The complete derivation of the formula for the covariance matrix of unknown parameters is too complex to be presented here. Therefore, just the formula is given

$$\mathbf{C}_{xx} = s_0^2 \cdot \mathbf{Q}_{xx} = s_0^2 \cdot (\mathbf{A}^{\mathrm{T}}\mathbf{P}\mathbf{A})^{-1}. \tag{2.16}$$

2.1.9 Positional Accuracy Improvement as an Adjustment Problem

In this section, we want to give a typical example of a PAI adjustment problem. First, we consider the workflow from map digitalization to improved global coordinates from the point of view of adjustment. Then, we simplify the example to a one-dimensional problem and reproduce the process with a concrete calculation. Figure 2.6 depicts the workflow.

The result of scanning a map is a raster image with a row–column coordinate system. The coordinates of specific points (building corners, boundary points, etc.) are determined in the raster system. On the basis of the scan resolution, the raster coordinates can be converted into metrical map coordinates. If the scan resolution is measured in dots per inch (dpi), then the converting form is

$$\boldsymbol{x}_{\text{map}} = \boldsymbol{x}_{\text{raster}} \cdot \frac{0.0254\,\text{m}}{\text{resolution [dpi]}}.$$

The standard deviations and, thereby, the variances of the digitized map coordinates are dependent on the scale factor and the quality of the underlying map. As a rule of thumb,

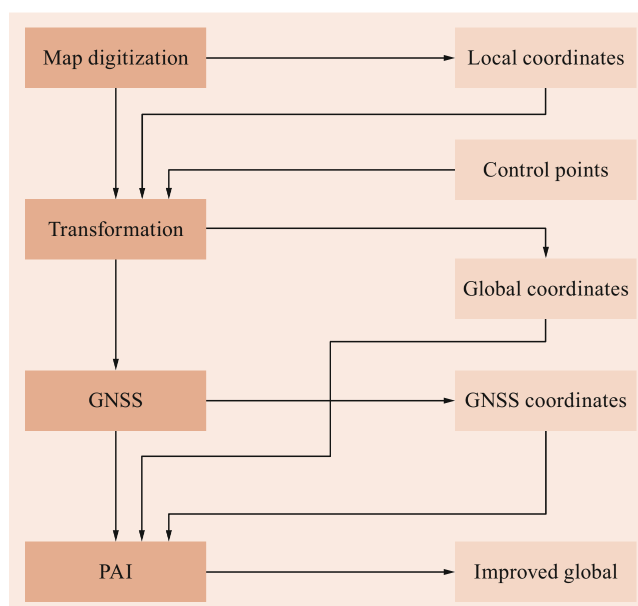
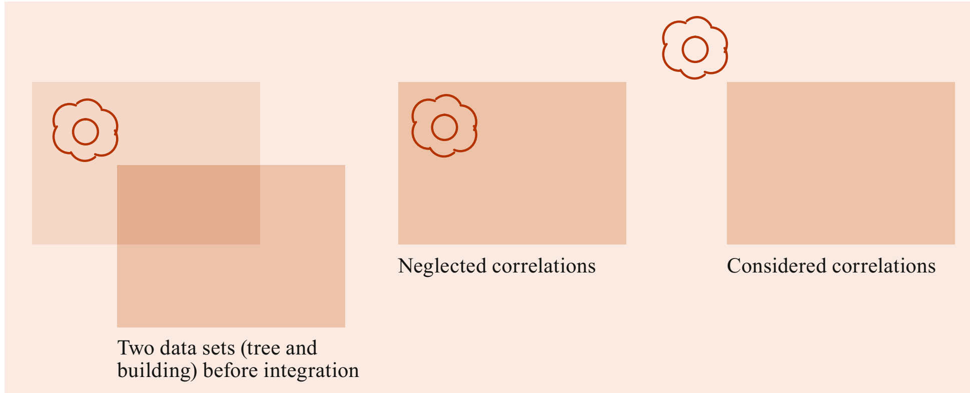


Fig. 2.6 PAI workflow for map digitization

Fig. 2.7 PAI neglecting and considering correlations



the standard deviation of a map coordinate value σ_{map} is about 0.5 mm.

The measured coordinates are random variables, and they are stochastically dependent because the point positions on a map result from real-world measurements and plots, both of which were made following the principle of adjacent points. Therefore, the coordinate values are stochastically dependent, analogous to the example given in Sect. 2.1.8 for adjustment with several unknown parameters.

However, in general, it is impossible to reconstruct a map history in detail, and therefore the covariances between digitized coordinates are not known. The outcome of a digitalization process is an observation vector $\boldsymbol{l}_r$ with map coordinates and a covariance matrix of these observations $\mathbf{C}_{ll}$ with known principal diagonal but unknown covariances.

Positional Accuracy Improvement

If the global coordinates of a GIS were stochastically independent, then we could just exchange the less accurate coordinates for more accurate ones. However, as they are stochastically dependent, such a procedure would lead to a violation of geometrical neighborhood relationships.

Figure 2.7 illustrates the necessity to consider stochastic dependencies between coordinates of adjacent points. In the example presented, coordinates with higher accuracy for the building corners were determined but not for the adjacent tree. If the correlations remain unconsidered, and the coordinates of the building are exchanged, then the tree seems to stand inside the building after PAI. However, if the stochastic dependencies are considered, the tree is shifted with the building during PAI.

2.1.10 Transformation

Map coordinates are transformed into a higher-level global reference frame. Control points with known coordinates in the local map coordinate system and the global reference frame are used to determine the necessary transformation parameters. The general transformation approach is

$$\boldsymbol{x}_{\text{global}} = \boldsymbol{t} + \mathbf{R} \cdot \boldsymbol{x}_{\text{map}}, \tag{2.17}$$

where $\boldsymbol{x}_{\text{global}}$ is the vector of global coordinates, $\boldsymbol{t}$ is the translation vector of transformation, $\mathbf{R}$ is the rotation matrix of transformation, and $\boldsymbol{x}_{\text{map}}$ is the vector of map coordinates.

As we can see, the global coordinates are linear functions of map coordinates. Applying the law of error propagation to (2.17) it is possible to calculate the standard deviations of global coordinates while their covariances remain unknown.

2.1.11 GNSS Measurement

For a number of points, global coordinates with higher accuracy are determined by GNSS measurements. Each GNSS point yields two redundant coordinates.

2.1.12 Improving Absolute Geometry

However, how can correlations be quantified and taken into account? One option is the introduction of artificial covariances. These can always be calculated as functions of the distances between two points. The smaller the distance between two points, the larger their covariance becomes. However, in practice, this approach is difficult to handle. Often there are more than 100,000 points to be processed, which would lead to extremely large covariance matrices.

A more practical solution is the introduction of pseudo-observations. In this approach, stochastic dependencies between GIS coordinates are modeled by coordinate differences. The basis of the determination of these pseudo observations is Delaunay triangulation (Fig. 2.8).

Before triangulation, the point positions are expressed uniquely by global coordinates. This parameterization is exchanged with pseudo observations generated from coordinate differences of points that are adjacent in the triangle network. This new parameterization is redundant but still consistent (Fig. 2.9).

Now the observation vector $\boldsymbol{l}_{\text{GNSS}}$ is extended with GNSS coordinates of higher accuracy (Fig. 2.10).

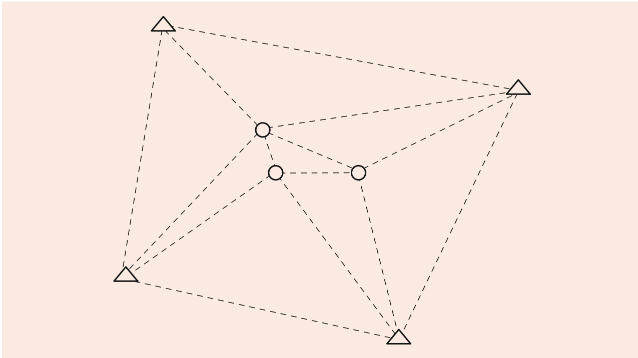


Fig. 2.8 Delaunay triangulation

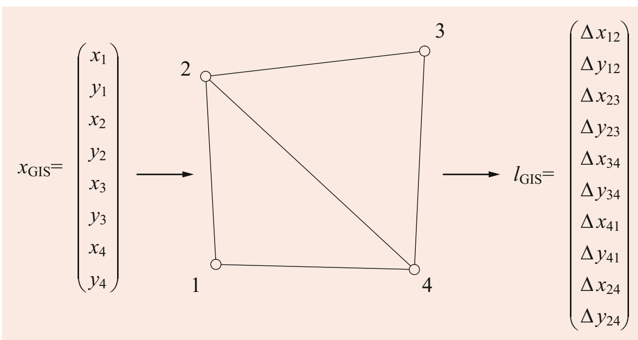


Fig. 2.9 Replacement of point coordinates by pseudo observations (distances)

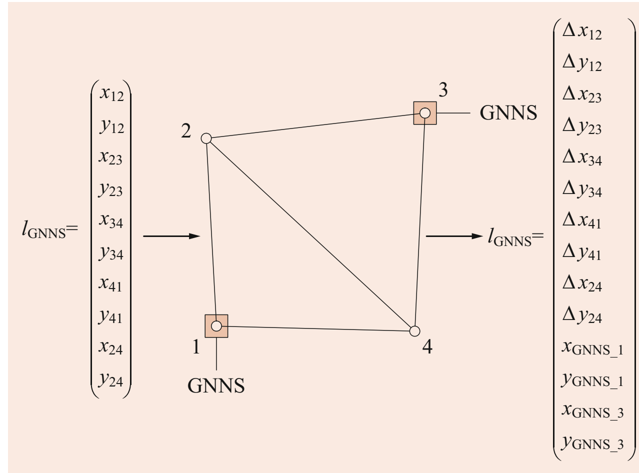


Fig. 2.10 Extension of the observation vector by GNSS coordinates

This leads to an adjustment problem with the improved global coordinates as unknown parameters and the coordinate distances and GNSS coordinates as observations. The observation equations have the structure

$$\begin{aligned} &\vdots \\ \Delta x_{ij} + v_{\Delta x} &= x_j - x_i\,, \\ \Delta y_{ij} + v_{\Delta y} &= y_j - y_i\,, \\ &\vdots \\ x_{GNSS\_i} + v_x &= x_i\,, \\ y_{GNSS\_i} + v_y &= y_i\,. \\ &\vdots \end{aligned}$$

These equations represent the functional model of the adjustment. The stochastic model is determined by the weights of the observations, which are functions of their standard deviations. The standard deviations of the GNSS coordinates depend on the measurement procedure applied and on the reproduction accuracy of the measured points. A practical value is $\sigma_{xy} = \pm 2\,\text{cm}$. The standard deviations of the coordinate differences are functions of those of the underlying map coordinates and of the corresponding point distances.

We want to explain this with an example. If map coordinates with standard deviation of $\sigma_{xy} = \pm 1\,\text{m}$ are given, then the formula for the calculation of the standard deviation of a coordinate difference could be

$$\sigma_\Delta = \sigma_{xy}\frac{d}{d_0} \quad \text{with} \quad d_0 = 100\,\text{m}.$$

In this case, an observational weight is inversely proportional to the square of the point distance.

The results of adjustment are improved coordinates for all points and their covariance matrix. The point accuracy depends, then, on the distance to the newly introduced GNSS points.

2.1.13 Improving Relative Geometry

Accuracy improvement of GIS coordinates is not exclusively effected by the introduction of precise global coordinates. There are also observations describing the relative geometry between adjacent points; for instance, it is known that, mostly, walls of buildings are rectangular and parcel limits are straight. Such geometrical constraints can be modeled by the introduction of corresponding observations. Rectangularity, for instance, can be expressed by a scalar product. The standard deviation of this scalar product observation can be calculated by error propagation of the corresponding point coordinates. Because construction workers can build a house with an accuracy of approximately 2 cm, a standard deviation of coordinates of about the same size would be adequate.

2.2 2-D Datum Transformations

The position of points can be defined in a variety of different local coordinate systems, e. g., in the instrument system of a total station, the local system of a digitized map, or as pixel coordinates of an orthophoto. In such a case, usually it is the objective of processing to determine the coordinates of these points in a global reference frame. On the other hand, datasets often use a different datum for point coordinates. In all these cases, it is necessary to transform point coordinates from one datum into another.

Coordinate transformation can generally be described as a mapping of a metric space R_s into another metric space R_t. The mapping is modeled by a transformation function $f(\boldsymbol{p})$, where $\boldsymbol{p}$ is the vector of the transformation parameters, i. e.,

$$f(p)\colon R_s \to R_t .$$

In the following, we call the space R_s the start system s, and R_t the target system t.

Frequently, the transformation parameters are not known a priori. Then, the parameters have to be determined by a calculation using control points. The coordinates of the control points are known in both the start and the target systems, whereas the coordinates of the new points are known only in the start system.

In the majority of cases, the number of control point coordinates exceeds the number of transformation parameters to be determined. In such cases, the calculation of transformation parameters is an adjustment problem.

In the general approach, the control point coordinates in both the start and the target system are considered as observations. This leads to an adjustment by a (mostly) nonlinear Gauss–Helmert model, in which the connection between observations and unknowns is expressed in an implicit way as

$$f[(\boldsymbol{l}+\boldsymbol{v}), p] = 0 \quad \text{with} \quad \boldsymbol{l} = \begin{pmatrix} \boldsymbol{x}_s \\ \boldsymbol{x}_t \end{pmatrix}. \tag{2.18}$$

For simplicity, we will use a Gauss–Markov model with the coordinates in the target system x_t as observations and the coordinates in the start system as constants, i. e.,

$$\boldsymbol{x}_t + \boldsymbol{v} = f(p, \boldsymbol{x}_s). \tag{2.19}$$

2.2.1 Centroid Reduction

For several reasons, it is appropriate to perform centroid reduction before the actual transformation calculation takes place. For the four and six-parameter approaches, centroid reduction leads to a diagonal structure of the normal equation matrix and simple closed forms for the calculation of the transformation parameters. For nonlinear approaches, this significantly improves the condition of the normal equation system. If the distance between the coordinate origin and coordinate centroid in the start system is much larger than the size of the transformed point set, a calculation may become impossible without centroid reduction.

First, the centroid coordinates of the control points of the start system are calculated as

$$x_c = \frac{1}{n}\sum_{i=1}^{n} x_{si}, \quad y_c = \frac{1}{n}\sum_{i=1}^{n} y_{si}. \tag{2.20}$$

Then, the centroid coordinates are subtracted from all (control points and new points) coordinates of the start system, thus

$$\begin{aligned} x'_{si} &= x_{si} - x_c, \\ y'_{si} &= y_{si} - y_c. \end{aligned} \tag{2.21}$$

The transformation parameters are calculated using the reduced coordinates of the control points as

$$p = f(\boldsymbol{x}'_s, \boldsymbol{x}_t). \tag{2.22}$$

With the now known transformation parameters, the coordinates of the new points in the target system can be calculated as

$$\boldsymbol{x}_t = f(p, \boldsymbol{x}'_s). \tag{2.23}$$

2.2.2 The Four-Parameter (Helmert) Transformation

The four-parameter transformation is the most common approach in GIS. It provides four degrees of freedom: two translations, one rotation, and one scale factor. The residual equations have the structure

$$\begin{aligned} x_{ti} + v_{xi} &= t_x + \cos\sigma \cdot s \cdot x'_{si} - \sin\varphi \cdot s \cdot y'_{si}, \\ y_{ti} + v_{yi} &= t_y + \sin\sigma \cdot s \cdot x'_{si} + \cos\varphi \cdot s \cdot y'_{si}, \end{aligned} \tag{2.24}$$

where t_x is the translation in x, t_y is the translation in y, φ is the rotation angle, and s is the scale factor.

Usually, the expressions $\cos\varphi \cdot s$ and $\sin\varphi \cdot s$ are substituted by the variables a and b, which act as unknowns in the adjustment calculation. We then get

$$\begin{aligned} x_{ti} + v_{xi} &= t_x + a \cdot x'_{si} - b \cdot y'_{si}, \\ y_{ti} + v_{yi} &= t_y + b \cdot x'_{si} + a \cdot y'_{si}. \end{aligned} \tag{2.25}$$

2

The original coordinates of the start system are substituted by their centroid-reduced coordinates, thus

$$\begin{aligned} x_{ti} + v_{xi} &= t'_x + a \cdot x'_{si} - b \cdot y'_{si}, \\ y_{ti} + v_{yi} &= t'_y + b \cdot x'_{si} + a \cdot y'_{si}. \end{aligned}$$

This approach can also be written in matrix notation as

$$\boldsymbol{x}_t + \boldsymbol{v} = \boldsymbol{t} + \mathbf{R}s \cdot \boldsymbol{x}'_s = \boldsymbol{t} + \mathbf{D} \cdot \boldsymbol{x}'_s \quad \text{with} \quad \boldsymbol{t} = \begin{pmatrix} t_x \\ t_y \end{pmatrix},$$

$$\mathbf{D} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} \cos\varphi \cdot s & -\sin\varphi \cdot s \\ \sin\varphi \cdot s & \cos\varphi \cdot s \end{pmatrix}.$$

For the calculation of the translation parameters, it is necessary to also determine the control point centroid coordinates in the target system. The parameters t_x and t_y can then be calculated directly as differences of the control point centroid coordinates between the start and the target system, i. e.,

$$\begin{aligned} t_x &= x_{tc} - x_{sc}, \\ t_y &= y_{tc} - y_{sc}. \end{aligned} \tag{2.26}$$

The substituted parameters a and b are calculated from the reduced control point coordinates in the start system and the control point coordinates in the target system as

$$\begin{aligned} a &= \frac{\sum\left(x'_{si} x_{ti} + y'_{si} y_{ti}\right)}{\left(x'^2_{si} + y'^2_{si}\right)}, \\ b &= \frac{\sum\left(x'_{si} y_{ti} + y'_{si} x_{ti}\right)}{\left(x'^2_{si} + y'^2_{si}\right)}. \end{aligned} \tag{2.27}$$

The target coordinates of the new points can now be calculated by inserting the reduced start coordinates into (2.25).

By converting (2.25) and inserting the start and target coordinates of the control points, we are able to calculate the residuals as

$$\begin{aligned} v_{xi} &= t_x + a \cdot x'_{si} - b \cdot y'_{si} - x_{ti}, \\ v_{yi} &= t_y + b \cdot x'_{si} + a \cdot y'_{si} - y_{ti}. \end{aligned} \tag{2.28}$$

A general value for the evaluation of the achieved accuracy is the empirical standard deviation $\hat{\sigma}_0$ of the observed target coordinates. This value can be calculated from the residuals as

$$\begin{aligned} \hat{\sigma}_0^2 &= \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}}{2n_{\mathrm{p}} - 4} = \frac{\sum\left(v_{xi}^2 + v_{yi}^2\right)}{2n_p - 4}, \\ \hat{\sigma}_0 &= \sqrt{\hat{\sigma}_0^2}, \end{aligned} \tag{2.29}$$

where n_{p} is the number of control points.

By variance propagation, the empirical variances of the transformation parameters can now be calculated as

$$\hat{\sigma}_{\mathrm{tx}}^2 = \hat{\sigma}_{\mathrm{ty}}^2 = \frac{\hat{\sigma}_0^2}{n_{\mathrm{p}}}, \tag{2.30}$$

$$\hat{\sigma}_{\mathrm{a}}^2 = \hat{\sigma}_{\mathrm{o}}^2 = \frac{\hat{\sigma}_0^2}{\sum\left(x'^2_s + y'^2_s\right)}. \tag{2.31}$$

With these values, the empirical variances (and thereby standard deviations) of the transformed new point coordinates in the target system can be calculated as

$$\begin{aligned} \hat{\sigma}_{xt}^2 &= \hat{\sigma}_{yt}^2 = \hat{\sigma}_{\mathrm{tx}}^2 + \hat{\sigma}_a^2 \cdot \left(x'^2_{\mathrm{s}} + y'^2_{\mathrm{s}}\right), \\ \hat{\sigma}_{xt} &= \hat{\sigma}_{yt} = \sqrt{\hat{\sigma}_{xt}^2}. \end{aligned} \tag{2.32}$$

The standard deviations of the transformed coordinates are dependent on the distance to the centroid in the start system. This is due to the influence of the uncertainty of the rotation parameter, which grows with this distance.

Sometimes, the rotation angle σ and the scale factor s are required, not just the substituted unknowns a and b. These values can be derived from a and b as

$$\begin{aligned} \sigma &= \arctan\frac{b}{a}, \\ s &= \sqrt{a^2 + b^2}. \end{aligned}$$

2.2.3 Six-Parameter (Affine) Transformation

The six-parameter transformation, like the four-parameter transformation, is a linear adjustment problem. Also here, the transformation parameters can be calculated by simple sum formulas. This approach provides six degrees of freedom: two translations, two rotations, and two scale factors. The residual equations have the following structure

$$\begin{aligned} x_{ti} + v_{\mathrm{xi}} &= t_x + \cos\sigma_x \cdot s_x \cdot x'_{si} - \sin\sigma_x \cdot s_x \cdot y'_{si}, \\ y_{ti} + v_{\mathrm{yi}} &= t_y + \sin\sigma_y \cdot s_y \cdot x'_{si} + \cos\sigma_y \cdot s_y \cdot y'_{si}, \end{aligned} \tag{2.33}$$

where t_x is the translation in x, t_y is the translation in y, σ_x is the rotation angle of the x-axis, σ_y is the rotation angle of the y-axis, s_x is the scale factor in x, and s_y is the scale factor in y.

The expressions $\cos\sigma_x \cdot s_x$, $\sin\sigma_x \cdot s_x$, $\cos\sigma_y \cdot s_y$, and $\sin\sigma_y \cdot s_y$ can be substituted by the variables a–d, which act as unknowns in the adjustment calculation. We then get

$$\begin{aligned} x_{ti} + v_{xi} &= t_x + a \cdot x'_{si} + b \cdot y'_{si}, \\ y_{ti} + v_{yi} &= t_y + c \cdot x'_{si} + d \cdot y'_{si}. \end{aligned} \tag{2.34}$$

This approach can also be written in matrix notation as

$$\boldsymbol{x}_\mathrm{t} + \boldsymbol{v} = \boldsymbol{t} + \mathbf{R}s \cdot \boldsymbol{x}'_\mathrm{s} = \boldsymbol{t} + \mathbf{D} \cdot \boldsymbol{x}'_\mathrm{s} \quad \text{with} \quad \boldsymbol{t} = \begin{pmatrix} t_x \\ t_y \end{pmatrix},$$
$$\mathbf{D} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \cos\sigma_x \cdot s_x & -\sin\sigma_x \cdot s_x \\ \sin\sigma_y \cdot s_y & \cos\sigma_y \cdot s_y \end{pmatrix}. \tag{2.35}$$

As in the four-parameter approach, the parameters t_x and t_y are directly calculated as differences of the control point centroid coordinates between the start and the target system as

$$t_x = x_\mathrm{tc} - x_\mathrm{sc},$$
$$t_y = y_\mathrm{tc} - y_\mathrm{sc}.$$

The substituted parameters a–d are calculated from the reduced control point coordinates in the start system and the control point coordinates in the target system using

$$a = \frac{\sum x'_{si} x_\mathrm{ti}}{x'^2_{si}}, \quad b = \frac{\sum y'_{si} x_\mathrm{ti}}{y'^2_{si}},$$
$$c = \frac{\sum x'_{si} y_\mathrm{ti}}{x'^2_{si}}, \quad d = \frac{\sum y'_{si} y_\mathrm{ti}}{y'^2_{si}}. \tag{2.36}$$

The target coordinates of the new points can now be calculated by inserting the reduced start coordinates into (2.36).

By converting (2.17) and inserting the start and target coordinates of the control points, we are able to calculate the residuals as

$$v_{xi} = t_x + a \cdot x'_{si} + b \cdot y'_{si} - x_{\mathrm{t}i},$$
$$v_{yi} = t_y + c \cdot x'_{si} + d \cdot y'_{si} - y_{\mathrm{t}i}. \tag{2.37}$$

The empirical standard deviation $\hat{\sigma}_0$ of the observed target coordinates can be calculated by (2.30), analogous to the four-parameter transformation.

Variance propagation provides the empirical variances of the transformation parameters as

$$\hat{\sigma}^2_{\mathrm{t}x} = \hat{\sigma}^2_{\mathrm{t}y} = \frac{\hat{\sigma}^2_0}{n_\mathrm{p}}, \tag{2.38}$$

$$\hat{\sigma}^2_a = \hat{\sigma}^2_c = \frac{\hat{\sigma}^2_0}{\sum x'^2_s}, \tag{2.39}$$

$$\hat{\sigma}^2_b = \hat{\sigma}^2_d = \frac{\hat{\sigma}^2_0}{\sum y'^2_s}. \tag{2.40}$$

With these values, the empirical variances (and thereby the standard deviations) of the transformed new point coordinates in the target system can be calculated as

$$\hat{\sigma}^2_{x\mathrm{t}} = \hat{\sigma}^2_{\mathrm{t}x} + \hat{\sigma}^2_a \cdot x'^2_\mathrm{s} + \hat{\sigma}^2_b \cdot y'^2_\mathrm{s}, \quad \hat{\sigma}_{x\mathrm{t}} = \sqrt{\hat{\sigma}^2_{x\mathrm{t}}}, \tag{2.41}$$

$$\hat{\sigma}^2_{y\mathrm{t}} = \hat{\sigma}^2_{\mathrm{t}y} + \hat{\sigma}^2_c \cdot x'^2_\mathrm{s} + \hat{\sigma}^2_d \cdot y'^2_\mathrm{s}, \quad \hat{\sigma}_{y\mathrm{t}} = \sqrt{\hat{\sigma}^2_{y\mathrm{t}}}. \tag{2.42}$$

Also here, the standard deviations of the transformed coordinates are dependent on the distance to the centroid in the start system.

The rotation angles φ_x and φ_y and the scale factors s_x and s_y can be derived from a–d as follows

$$\varphi_x = \arctan\frac{b}{a}, \qquad \varphi_y = \arctan\frac{d}{c}, \tag{2.43}$$

$$s_x = \sqrt{a^2 + b^2}, \qquad s_y = \sqrt{c^2 + d^2}. \tag{2.44}$$

2.2.4 Three-Parameter Transformation

Unlike the four and six-parameter approaches, the three-parameter transformation is a nonlinear adjustment problem. The solution is found through an iterative process by applying Newton's method. At the start of the iteration process, approximate values for the transformation parameters are needed. These approximate values can easily be obtained by performing a linear four-parameter transformation.

The three degrees of freedom here are two translations and one rotation. A scale factor is not modeled. The residual equations have the structure

$$x_{\mathrm{t}i} + v_{xi} = t_x + \cos\varphi \cdot x'_{\mathrm{s}i} - \sin\varphi \cdot y'_{\mathrm{s}i},$$
$$y_{\mathrm{t}i} + v_{yi} = t_y + \sin\varphi \cdot x'_{\mathrm{s}i} + \cos\varphi \cdot y'_{\mathrm{s}i}. \tag{2.45}$$

In general, we can say that the vector of the adjusted control point coordinates is a function of the parameter vector (the coordinates in the start systems are constants), i. e.,

$$\boldsymbol{x}_\mathrm{t} + \boldsymbol{v} = f(p). \tag{2.46}$$

In the case considered, $f(\boldsymbol{p})$ is a nonlinear function, and, therefore, $f(\boldsymbol{p})$ has to be linearized. The linearization is done by developing $f(\boldsymbol{p})$ in a Taylor series of degree 1 as

$$\boldsymbol{x}_\mathrm{t} + \boldsymbol{v} = f(p_0) + \frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} \cdot \underbrace{(p - p_0)}_{\Delta p}. \tag{2.47}$$

In this expression, $\boldsymbol{p}_0$ is the vector of the proximity parameters. The vector $\Delta\boldsymbol{p}$ contains the substitute unknowns. To get the substitute observations, we rearrange expression (2.47) to

$$\underbrace{\boldsymbol{x}_\mathrm{t} - f(p_0)}_{\tilde{x}} + \boldsymbol{v} = \frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} \cdot \Delta p. \tag{2.48}$$

In this form, $\tilde{\boldsymbol{x}}$ is the vector of the substitute observations. Its elements are

$$\tilde{x}_{\mathrm{t}i} = x_{\mathrm{t}i} - t^0_x - \cos\varphi^0 \cdot x'_{si} + \sin\varphi^0 \cdot y'_{si},$$
$$\tilde{y}_{\mathrm{t}i} = y_{\mathrm{t}i} - t^0_y - \sin\varphi^0 \cdot x'_{si} - \cos\varphi^0 \cdot y'_{si}. \tag{2.49}$$

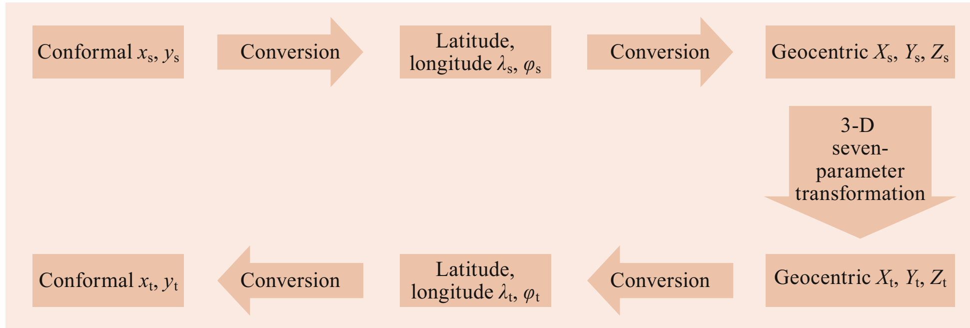


Fig. 2.11 Workflow of a 3-D seven-parameter transformation

The linearized residual equations are then

$$\tilde{x}_{ti} + v_{xi} = \Delta t_x - (\sin\varphi^0 \cdot x'_{si} + \cos\varphi^0 \cdot y'_{si}) \cdot \Delta\varphi,$$
$$\tilde{y}_{ti} + v_{yi} = \Delta t_y + (\cos\varphi^0 \cdot x'_{si} - \sin\varphi^0 \cdot y'_{si}) \cdot \Delta\varphi. \quad (2.50)$$

The overdetermined equation system (2.50) can now be solved by using the rules of nonlinear Gauss–Markov model adjustment.

2.2.5 Five-Parameter Transformation

Like the three-parameter transformation, the five-parameter transformation is a nonlinear adjustment problem. The degrees of freedom are two translations, one rotation and two scale factors. The nonlinear residual equations are

$$x_{ti} + v_{xi} = t_x + (\cos\varphi \cdot x'_{si} - \sin\varphi \cdot y'_{si}) \cdot s_x,$$
$$y_{ti} + v_{yi} = t_y + (\sin\varphi \cdot x'_{si} + \cos\varphi \cdot y'_{si}) \cdot s_y. \quad (2.51)$$

The equations for the substitute observations are

$$\tilde{x}_{ti} = x_{ti} - t_x^0 - (\cos\varphi^0 \cdot x'_{si} + \sin\varphi^0 \cdot y'_{si}) \cdot s_x^0,$$
$$\tilde{y}_{ti} = y_{ti} - t_y^0 - (\sin\varphi^0 \cdot x'_{si} - \cos\varphi^0 \cdot y'_{si}) \cdot s_y^0. \quad (2.52)$$

The linearized residual equations are

$$\tilde{x}_{ti} + v_{xi} = \Delta t_x - (\sin\varphi^0 \cdot x'_{si} + \cos\varphi^0 \cdot y'_{si}) \cdot s_x^0 \cdot \Delta\varphi$$
$$+ (\cos\varphi^0 \cdot x'_{si} - \sin\varphi^0 \cdot y'_{si}) \cdot \Delta s_x,$$
$$\tilde{y}_{ti} + v_{yi} = \Delta t_y + (\cos\varphi^0 \cdot x'_{si} - \sin\varphi^0 \cdot y'_{si}) \cdot s_y^0 \cdot \Delta\varphi$$
$$+ (\sin\varphi^0 \cdot x'_{si} + \cos\varphi^0 \cdot y'_{si}) \cdot \Delta s_y. \quad (2.53)$$

Like the three-parameter transformation, the overdetermined system of the linearized residual equations (2.53) can be solved with Gauss–Markov model adjustment.

2.2.6 Conformal Transformation with Complex Polynomials

A map is called conformal (or angle preserving) if it preserves oriented angles between curves with respect to their orientation (i. e., not just the acute angle). Conformal maps preserve both angles and the shapes of infinitesimally small figures, but not necessarily their size [3, Chap. 8].

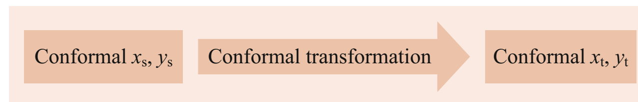


Fig. 2.12 Workflow of a 2-D conformal transformation

Many projections used in GIS are conformal mappings of a double-curved reference surface (ellipsoid or sphere) into a plane. Examples are the Universal Transverse Mercator (UTM), Gauss–Krüger, stereographic, or Lambert projection. Frequently, the task is a datum transformation of point sets whose coordinates are given as a conformal projection. The classical approach for the solution of this task is a three-dimensional (3-D), seven-parameter transformation (Fig. 2.11). For this purpose, the projected coordinates have to be converted into 3-D geocentric Cartesian coordinates. A prerequisite for that conversion is the knowledge of the ellipsoidal heights of the control points in both the start and the target systems. Figure 2.11 shows the workflow of this transformation.

An alternative to this lengthy approach can be a 2-D conformal transformation (Fig. 2.12). This approach directly maps the conformal coordinates x_s, y_s of the start system into the coordinates x_t, y_t in the target system.

The underlying idea of this approach is the mathematical rule that the analytical function of a complex variable is a conformal mapping. One class of analytical functions is the polynomials. This means that a Taylor series developing a complex variable always results in a conformal mapping

$$z_t = \alpha_0 + \alpha_1 \cdot z_s + \alpha_2 \cdot z_s^2 + \alpha_3 \cdot z_s^3 + \ldots + \alpha_n \cdot z_s^n. \quad (2.54)$$

We express the coordinates of our control points x_s, y_s, x_t, and y_t as complex numbers

$$z'_s = x'_s + y'_s i,$$
$$z_t = x_t + y_t i. \quad (2.55)$$

Then, we develop z_t as a polynomial function of z'_s of degree 1, i. e.,

$$z_t = \alpha_0 + \alpha_1 \cdot z_s. \quad (2.56)$$

If we expand expression (2.54), then we get

$$\begin{aligned} x_t + y_t i &= \mathrm{Re}(\alpha_0) + \mathrm{Im}(\alpha_0)i \\ &\quad + [\mathrm{Re}(\alpha_1) + \mathrm{Im}(\alpha_1)i] \cdot (x'_s + y'_s i) \\ &= \mathrm{Re}(\alpha_0) + \mathrm{Im}(\alpha_0)i \\ &\quad + \mathrm{Re}(\alpha_1) \cdot x'_s - \mathrm{Im}(\alpha_1) \cdot y'_s \\ &\quad + [\mathrm{Im}(\alpha_1) \cdot x'_s + \mathrm{Re}(\alpha_1) \cdot y'_s]i . \end{aligned} \tag{2.57}$$

After separation of the real and imaginary parts we get the equations for the coordinates x_t and y_t as

$$\begin{aligned} x_t &= \mathrm{Re}(\alpha_0) + \mathrm{Re}(\alpha_1) \cdot x'_s - \mathrm{Im}(\alpha_1) \cdot y' , \\ y_t &= \mathrm{Im}(\alpha_0) + \mathrm{Im}(\alpha_1) \cdot x'_s + \mathrm{Re}(\alpha_1) \cdot y' . \end{aligned} \tag{2.58}$$

If we compare (2.58) with (2.8), we see that the four-parameter transformation is just a special case of complex transformation polynomials, namely the case of degree 1. The components of α_0 are identical to the components of the translation vector $\boldsymbol{t}$, whereas the components of α_1 are identical to a and b.

In the four-parameter approach, the rotation and scale factor are constant over the whole area under consideration. However, if we transform large areas, e. g., the size of a country, this assumption is no longer valid. In that case, the rotation angle and scale factor are functions of the point coordinates. To be able to model this property it is necessary to extend the complex transformation polynomial to elements of higher order.

The necessary degree of development depends on several factors, especially on the size of the area. The calculated elements of the polynomial can be tested for significance by using Student's t-test [4].

The adjustment approach is presented here for the example of a polynomial of degree 2. The extension to higher degrees works analogously. The residual equations have the structure

$$\begin{aligned} x_{ti} + v_{xi} &= \mathrm{Re}(\alpha_0) + \mathrm{Re}(\alpha_1) \cdot x'_s \\ &\quad - \mathrm{Im}(\alpha_1) \cdot y' + \mathrm{Re}(\alpha_2) \cdot \mathrm{Re}(z_s'^2) \\ &\quad - \mathrm{Im}(\alpha_2) \cdot \mathrm{Im}(z_s'^2) + \ldots , \\ y_t + v_{yi} &= \mathrm{Im}(\alpha_0) + \mathrm{Im}(\alpha_1) \cdot x'_s \\ &\quad + \mathrm{Re}(\alpha_1) \cdot y' + \mathrm{Im}(\alpha_2) \cdot \mathrm{Re}(z_s'^2) \\ &\quad + \mathrm{Re}(\alpha_2) \cdot \mathrm{Im}(z_s'^2) + \ldots \end{aligned} \tag{2.59}$$

As we can see, the equations are linear. The coefficients are the real and imaginary parts of the powers of z'_s. The unknowns are the real and imaginary parts of the values α_i. The unknowns can be calculated by a linear Gauss–Markov adjustment calculation.

2.2.7 Modeling of Correlations

A special problem in the integration of spatial data is distance-dependent correlations among coordinates within one dataset. This section discusses the reason for these correlations and shows how they can be modeled in an adjustment approach. It is shown that techniques such as proximity fitting, rubber sheeting, and constraint management can be traced back to the same adjustment problem.

2.2.8 Reasons for Correlations

In the pre-GNSS era, measurements were done following the principle of neighborhoods. Trigonometric networks were calculated from large to small. Manual mapping was done in the same manner. These circumstances led to the fact that coordinates originating from digitized paper maps or classical terrestrial surveys show distance-dependent correlations. In other words, one can say that the relative accuracy of two neighboring points is higher than their absolute accuracy relative to the reference frame. This fact can be expressed in the following equation by the covariance of two coordinate values

$$\begin{aligned} \Delta x &= x_B - x_A , \\ \sigma_{\Delta x} &= \sqrt{\sigma^2_{\Delta x_A} + \sigma^2_{\Delta x_B} - 2 \cdot \mathrm{cov}(x_A, x_B)} \\ &\quad \text{with} \quad \mathrm{cov}(x_A, x_B) \neq 0 . \end{aligned} \tag{2.60}$$

The quotient of the covariance $\mathrm{cov}(x_A, x_E)$ and the standard deviations of the coordinates σ_A and σ_E represent the correlation coefficient of the coordinate values x_A and x_E, i. e.,

$$\rho_{x_A, x_B} = \frac{\mathrm{cov}(x_A, x_B)}{\sigma_A \cdot \sigma_B} . \tag{2.61}$$

The correlation coefficient always has a value between -1 and 1. In the case of the coordinates considered, ρ_{x_A, x_B} is a function of the distance between points A and B with a codomain between 1 and 0 (Fig. 2.13).

As we can see, the correlation grows as the distance gets smaller, and it converges towards zero as the distance gets longer.

Many of the problems regarding integration of different GIS datasets are caused by these distance-dependent correlations. If they did not exist, all discrepancies accruing in the control points after a datum transformation could be considered as random errors. It would be enough to average them, whereas all new points could keep their transformed coordinates.

However, because of the existence of distance-dependent correlations, it is necessary to account for them during the integration process. A simple transformation approach is

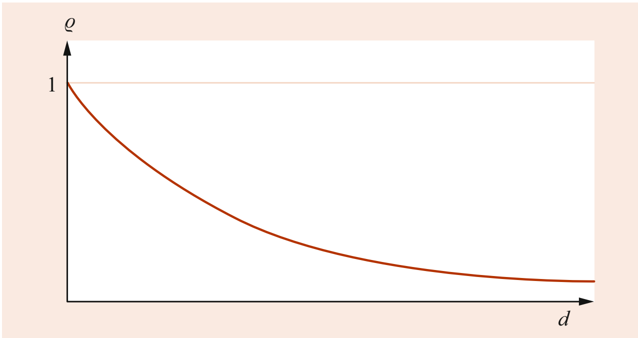


Fig. 2.13 Correlation coefficient as a function of distance

not adequate in many cases. The mapping model has to be extended, otherwise local geometrical relationships of neighboring points would be violated.

Several approaches exist to model these correlations. Some of them are presented and discussed in the following.

Because the genesis of coordinates can almost never be completely reconstructed, all modeling approaches are based on hypotheses. For this reason, there is never only one possible approach. It has to be decided for the individual case which model is appropriate.

2.2.9 Rubber Sheeting

Several GIS provide rubber-sheeting tools for conflation of different vector datasets. The basic idea of most of these tools is a six-parameter affine transformation applied on single triangles of a triangulated irregular network (TIN). The triangles are the result of Delaunay triangulation over the control points in the start system. The six-parameter approach provides, for three control points, a unique solution. For this reason, no discrepancies occur in the control points. The result is an individual set of transformation parameters for each triangle in the network. All new points lying in a particular triangle are subsequently transformed using the parameters belonging to that triangle (Fig. 2.14).

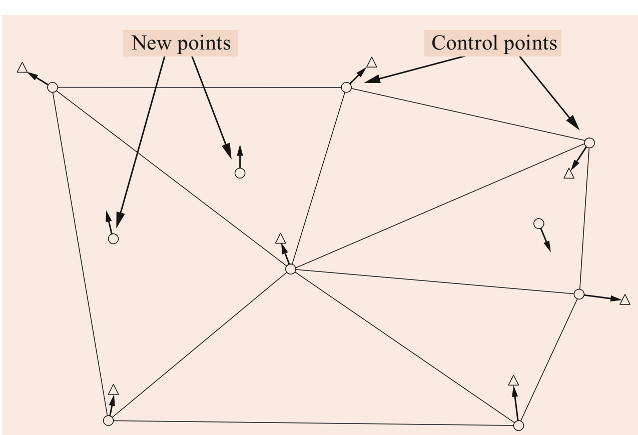


Fig. 2.14 Principle of rubber sheeting

In many cases, this model is adequate, but it also has disadvantages. The coordinates of the target system are considered to be correct. So, it is not possible to conflate two datasets where the coordinates in the start and the target systems are of limited accuracy. Because of the unambiguousness of the solution, it is not possible to detect blunders. Blunders are directly propagated to the new point coordinates. Also, geometrical constraints such as collinearities, rectangularities, or parallelisms are not retained.

2.2.10 Stochastic Modeling

A further option to model distance-dependent correlations is their direct calculation with a hypothetic function $\rho(d)$. Examples for such a function are

$$\rho(d) = \frac{1}{1 + \frac{d}{d_0}} \quad \text{or} \quad \rho(d) = \frac{1}{1 + \left(\frac{d}{d_0}\right)^2} . \tag{2.62}$$

In these functions, d is the distance, and d_0 is a constant value. Clearly, $\rho(0) = 1$ and $\rho(\infty) = 0$ apply. The hypothetic function as well as the correlation and the standard deviations of the point coordinates lead to the covariances of the coordinates

$$\operatorname{cov}(x_{\mathrm{A}}, x_{\mathrm{B}}) = \sigma_{x_{\mathrm{A}}} \cdot \sigma_{x_{\mathrm{B}}} \cdot \rho_{\mathrm{A,B}} . \tag{2.63}$$

The covariances are collated in the covariance matrix $\mathbf{C}_{xx}$. This covariance matrix is introduced in an adjustment process for the calculation of the transformation parameters and is also used for the propagation of the residuals to the coordinates of the new points. During an adjustment calculation, $\mathbf{C}_{xx}$ has to be inverted. However, in practice, the matrix $\mathbf{C}_{xx}$ can become very large, which leads to an enormous calculation effort. For this reason, the stochastic modeling approach is more of a theoretical proposal than a method that is applicable in practice.

2.2.11 Functional Modeling

Distance-dependent correlations originate from observations between neighboring points. In most cases, the original observations are not known, or their acquisition would require an indefensible effort. These problems can be solved by the introduction of artificial observations.

The introduction of these artificial observations is done after the datum transformation. The results of the datum transformation are the coordinates of all points in the target system as well as the coordinate residuals for the control points. Those residuals are caused by random errors and

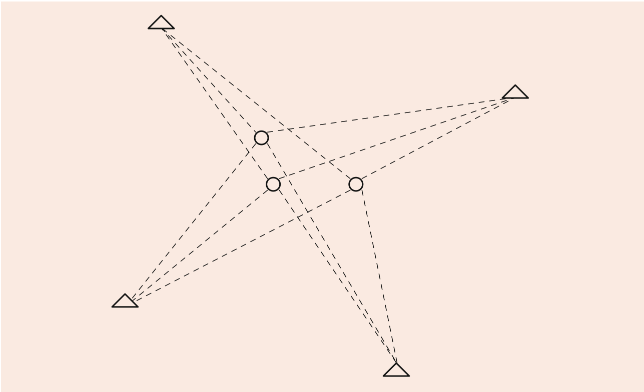


Fig. 2.15 Observation topology of distance-dependent averaging

distance-dependent correlations. To be able to model the correlations, the transformation approach is now extended by a further step: proximity fitting. For this purpose, artificial observations are introduced between points of the start system.

An easy approach is the distance-dependent distribution of coordinate residuals. For each new point, a weighted average of coordinate residuals is calculated and applied to its coordinates. The particular weights for the average calculation are functions of the distance between the new point and the neighboring control points, i. e.,

$$p_i = \frac{d_0}{d_i} \quad \text{or} \quad p_i = \left(\frac{d_0}{d_i}\right)^2 \quad \text{or} \quad p_i = \left(\frac{d_0}{d_i}\right)^{\frac{3}{2}} . \tag{2.64}$$

The result is identical to that of an adjustment calculation with coordinate difference observations between the new and the control points. It is advantageous that the adjustment for each new point can be calculated separately, which leads to very small calculation effort (Fig. 2.15).

A disadvantage of this approach is the absence of a direct link between the topology of observations and the real neighborhood relationship of points. The method fails if additional observations are introduced into the adjustment model. Such configurations may lead to significant violations of local geometry relationships.

Figure 2.16 shows the situation before and after the introduction of rectangularity observations in the adjustment. The rectangularity observations result in a change of the building geometry, while the tree in the lower-left corner keeps its position. The reason for this behavior is the nonmodeled neighborhood between the building corner and the tree. As a result, the tree is situated inside the house.

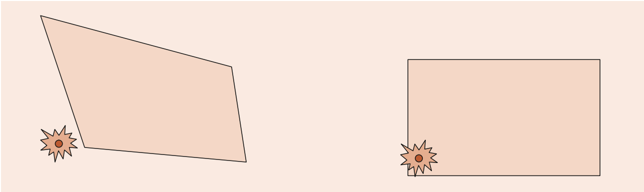


Fig. 2.16 Before and after the application of rectangularity constraints

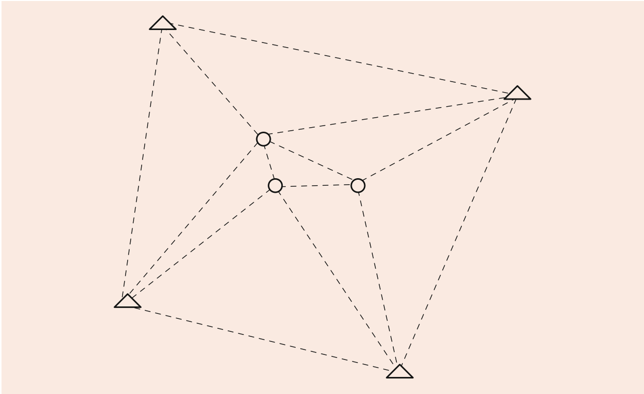


Fig. 2.17 Topology of observations (edges) of a Delaunay triangulation

A better way of creating an observation topology is Delaunay triangulation of all points (control and new points) of the considered dataset. The triangle edges are the carriers of neighborhood information. Along the triangle edges, artificial coordinate distance observations are generated (Fig. 2.17).

In this approach, the neighborhood relationships are modeled directly. Further observations, such as geometrical constraints, can be introduced into the adjustment without problems. It is disadvantageous that all point coordinates have to be introduced as unknowns in the same adjustment, which can lead to a very large normal equation matrix. Therefore, the solution of sophisticated geometrical data-integration problems requires special software.

2.2.12 Modeling of Point Identities

There are two methods to model point identities, in which the identity information is expressed either topologically or geometrically. The most common method is topological modeling of identities by introducing the same point identifier to corresponding points in different datasets. However, this method can lead to problems because of inevitable point confusions. Confusions only indirectly affect the residuals of corresponding observations. To eliminate such confusion, it is necessary to *diffuse* points by the generation of new point objects, whereby referential integrity requirements have to be observed.

Geometrical modeling of point identities is an alternative approach. An identity observation is introduced. This means that a coordinate difference with value zero is observed between two potentially identical points instead of assigning a common identifier. The square sum of the residuals of such

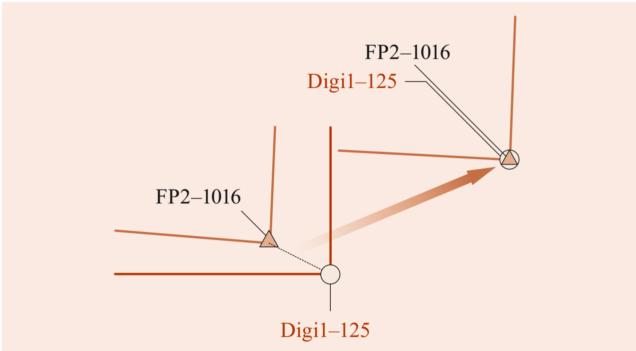


Fig. 2.18 Handling of point identities

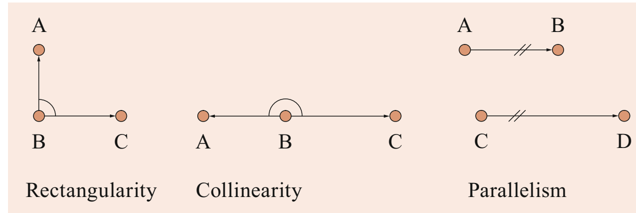


Fig. 2.19 Geometrical constraints

an identity observation is χ^2 distributed and can be tested for significance. Misidentifications can easily be detected and then eliminated.

Figure 2.18 shows the principle of a point identity observation. It is a relative measurement between two points with different point numbers. The point identity is weighted with a standard deviation derived from connected observations (e. g., map accuracy). It reacts like a *rubber band* (see the left-hand part of Fig. 2.18) with elasticity corresponding to its weight and can be analyzed like all other observations. Unreliable measurements can easily be removed without violating the topology. If all remaining point identities are reliable, their standard deviation is fixed (set to zero), and the connected points get the same coordinates (see the right-hand part of the figure). Finally, the points can be fused by a GIS to obtain a topology without redundancies.

2.2.13 Geometrical Constraints, Known Relative Geometry

Sometimes, conflation of the coordinates of different datasets is not the only task. Additionally, geometrical constraints may have to be retained, or known values such as distances, aligning bases, and so on, be considered.

Typical geometrical constraints are the rectangularity of buildings, the collinearity of boundary points, or the parallelism of waysides.

All these constraints can be expressed by either a scalar product or a cross product of two vectors. The vector components in this case are the coordinate differences of the points involved in the target system, i. e.,

$$\boldsymbol{v}_{\mathrm{AB}} = \begin{pmatrix} \Delta x_{\mathrm{AB}} \\ \Delta y_{\mathrm{AB}} \end{pmatrix} = \begin{pmatrix} x_{\mathrm{B}} - x_{\mathrm{A}} \\ y_{\mathrm{B}} - y_{\mathrm{A}} \end{pmatrix}, \qquad (2.65)$$

where (Fig. 2.19) rectangularity is expressed as

$$\boldsymbol{v}_{\mathrm{AB}} \cdot \boldsymbol{v}_{\mathrm{BC}} \overset{!}{=} 0, \qquad (2.66)$$

collinearity as

$$|\boldsymbol{v}_{\mathrm{AB}} \times \boldsymbol{v}_{\mathrm{BC}}| \overset{!}{=} 0, \qquad (2.67)$$

and parallelism as

$$|\boldsymbol{v}_{\mathrm{AB}} \times \boldsymbol{v}_{\mathrm{CD}}| \overset{!}{=} 0. \qquad (2.68)$$

In theory, there are two options to model these constraints in the adjustment approach. The first option is the formulation of restriction equations of unknowns. These equations force strict compliance with the constraints. However, this method has two disadvantages. Firstly, the normal equation matrix grows with each constraint by one row or column, which leads to enormous inflation of the calculation. Secondly, it is not possible to check particular constraints for plausibility. Constraints are often detected by automated snooping algorithms that are not completely error free. Therefore, it is important to detect erroneously identified constraints.

A better approach is the formulation of constraints as observations. The observation is in all cases zero. As an example, the residual equation for a rectangularity constraint is given by

$$0 + v = (x_{\mathrm{A}} - x_{\mathrm{B}})(x_{\mathrm{C}} - x_{\mathrm{B}}) + (y_{\mathrm{A}} - y_{\mathrm{B}})(y_{\mathrm{C}} - y_{\mathrm{B}}). \qquad (2.69)$$

This approach does not influence the dimension of the normal equation matrix. On the other hand, the normalized residual of each observation can be used as a test value for an outlier test.

The weight for the constraint observation is calculated by variance propagation from the coordinates to the observation value.

Analogously to the constraint observations, known measurements, such as lengths and widths of buildings, widths of streets, distances between objects, and so on, can be introduced into the adjustment calculation.

2.2.14 Matching and Constraint Snooping

A prerequisite for application of adjustment techniques for spatial data integration is knowledge about identical objects in the relevant datasets. This section describes how identical objects in different datasets can be automatically detected. It will be shown how identity information can be modeled by identity observations. The basics of statistical test theory will be given. Furthermore, it will be shown that matching and adjustment interact in an alternating iterative process.

2.2.15 Topology and Extraction of Subgraphs

Each vector dataset can be considered as a graph. According to the general definition, a graph is a special case of a topology, containing a set of two-valued subsets, called edges, defined on a basic set of vertices [5]. The graph itself contains no geometric information at all. Even if the particular vertices have geometrical properties (coordinates), the topological information of the graph is invariant to transformations of this geometry. Prerequisites for this behavior are a unique object identifier for each vertex and nonredundant storage of the geometry in the data model.

However, in most GIS, a vertex is identified by its coordinates and stored redundantly in the geometry of several lines or shapes. To be able to use graph operators it is necessary to reconstruct the topological information from such geometry.

Graph operators are helpful tools to find identical objects in different vector datasets. Each object in a vector dataset can be seen as a subgraph. A special case of a subgraph is a single vertex. Simple matching operators work on the basis of vertex identities. However, for complex vector datasets, the results of such operators are often ambiguous. To get unique results, it is necessary to use subgraphs of higher complexity. The higher the complexity of the subgraphs used, the lower the probability of ambiguities in the matching results. Figure 2.20 shows some examples of subgraphs.

With graph operators, subgraphs of the same type can be extracted from the two datasets to be matched. In a second step, their geometrical properties are compared and tested for identity. A very efficient approach for extraction of subgraphs with defined structure is the use of adjacency tensors. These tensors result from a sequence of tensor products of the graph's adjacency matrix with itself.

2.2.16 Geometrical Parameterization of Subgraphs

Identified subgraphs contain no geometrical information. However, to be able to test pairs of them for identity, it is necessary to quantify their geometrical properties. In general, one can distinguish between datum-dependent and datum-independent parameterization of their geometry.

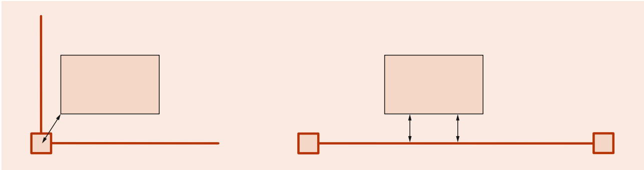


Fig. 2.21 Matching of corners and edges

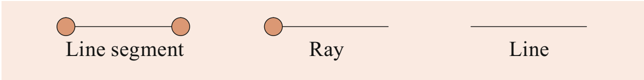


Fig. 2.22 A linear point set

First we consider the datum-dependent approach. A prerequisite for datum-dependent matching is that the vector datasets to be compared have to be in the same datum. Datum-dependent parameterization uses the coordinates of the vertices involved, or functions of them. The simplest approach is the direct use of coordinates. Using a vertex subgraph, one pair of coordinate tuples has to be compared; at an edge, two tuples have to be compared; at a corner, three tuples; and so on.

However, often a coordinate comparison is not appropriate, as illustrated in Fig. 2.21. If we compare the corner of the building with the boundary corner, we find that not all coordinate tuples of the involved vertices of the corners are identical. The same geometric fault is valid for the comparison of the building edge with the boundary edge.

In geometry, a linear point set can be given as a line segment, ray, or line (Fig. 2.22). The geometrical parameterization of these features is different. A line segment is simply parameterized by the coordinates of its end points

$$\boldsymbol{g}_{\text{line segment}} = \begin{pmatrix} x_{\text{A}} & y_{\text{A}} & x_{\text{B}} & y_{\text{B}} \end{pmatrix}^{\text{T}}, \tag{2.70}$$

where $\boldsymbol{g}_{\text{line segment}}$ is the vector of the geometrical parameters. A ray can be parameterized by the coordinates of its start point and a normalized direction vector as

$$\boldsymbol{g}_{\text{ray}} = \begin{pmatrix} x & y & d_x & d_y \end{pmatrix}^{\text{T}}. \tag{2.71}$$

Note that the covariance matrix of $\boldsymbol{g}_{\text{ray}}$ is singular because of the overparameterization of the direction. However, this overparameterization provides the advantage that the direction can be expressed without discontinuities. A line can be parameterized by its normal form $\boldsymbol{n}\boldsymbol{x} - d = 0$, where $\boldsymbol{n}$ is the normal vector of the line, $\boldsymbol{x}$ is the position vector of a point of the line, and d is the orthogonal distance between the line and the coordinate origin. Then, the parameter vector $\boldsymbol{g}_{\text{line}}$

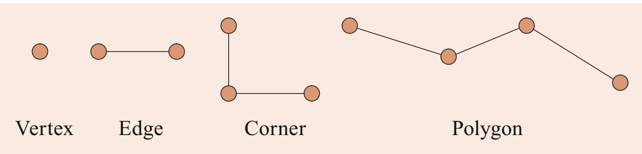


Fig. 2.20 Examples of subgraphs

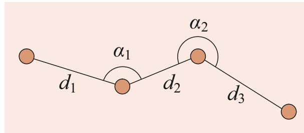


Fig. 2.23 Datum-independent parameterization of a polygon

contains the components of the normal vector and the translation parameter d, i. e.,

$$\boldsymbol{g}_{\text{line}} = \begin{pmatrix} n_x & n_y & d \end{pmatrix}^{\text{T}}. \tag{2.72}$$

Analogously to the ray exposition, $\boldsymbol{g}_{\text{line}}$ is overparameterized, and its covariance matrix is singular.

Subgraphs of different types can be parameterized by a combination of line segment, ray, and line parameterization. Each corner in Fig. 2.23 would be parameterized by one point and two normalized direction vectors as

$$\boldsymbol{g}_{\text{corner}} = \begin{pmatrix} x & y & d_{1x} & d_{1y} & d_{2x} & d_{2y} \end{pmatrix}^{\text{T}}. \tag{2.73}$$

The search for control points that are necessary for a transformation leads to the identification of identical objects in different vector datasets, which are not referenced to the same coordinate reference system. In such cases, the parameterization of the subgraphs must be done independently from the coordinate reference system. Let us take a closer look at the example of the polygon in Fig. 2.23. The inner geometry of this subgraph can be described by two angles and three distances.

Then, the parameter vector has the form

$$\boldsymbol{g}_{\text{polygon}} = \begin{pmatrix} \alpha_1 & \alpha_2 & d_1 & d_2 & d_3 \end{pmatrix}^{\text{T}}. \tag{2.74}$$

2.2.17 Search for Candidates

The aim of matching is to find subgraphs with identical geometrical parameter vectors in different vector datasets. However, if we think of corners, for instance, the number of extracted subgraphs can get very large.

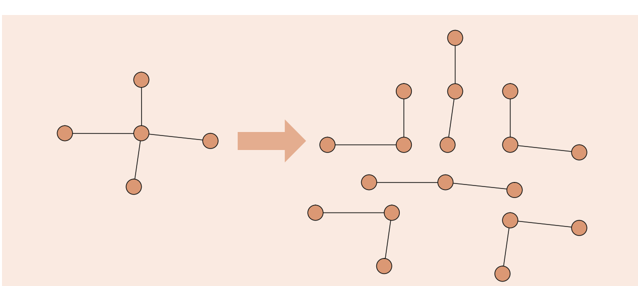


Fig. 2.24 Extracted corners

Figure 2.24 demonstrates that a simple junction of four edges leads to six corners. The total number of corners in a dataset can easily reach several hundred thousand.

If each parameter vector of one dataset is compared with each parameter vector of the other dataset, the number of comparisons increases with the square of the number of parameter vectors, i. e.,

$$\text{comparisons} = n\frac{n-1}{2}.$$

So, 100,000 corners require about 5×10^9 comparisons. With 500,000 corners, the number would reach about 10^{11}. Such a procedure results in a unacceptably long computing time.

An appropriate approach is the application of kd-trees. In the case of a corner with six geometrical parameters, a 6d search tree is created. In this tree, a search is performed in a k-dimensional window. In a well-balanced tree, the number of compare operations is the logarithmic function

$$\text{comparisons} \approx \log_d n.$$

Using a kd-tree, 100,000 corners require only seven comparisons, while the number reaches eight with about 500,000 corners (Chap. 3).

In any case, a kd window search results in a set of matching candidates. Whether one or more of them are identical in the start parameter vector can only be decided by the application of a statistical test.

2.2.18 Statistical Tests

To be able to test the parameter vectors $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ for identity, it is necessary to know their stochastic properties. The elements of $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ are random values. The elements inside one of these vectors are algebraically correlated because they are functions of the same set of random values—the coordinates of the vertices involved. The stochastic properties of the parameter vector elements are quantified by their covariance matrices $\mathbf{C}_{gg_1}$ and $\mathbf{C}_{gg_2}$.

These covariance matrices can be calculated by variance propagation from the coordinates to the parameters. First, the parameters have to be expressed as linear functions of the coordinates, i. e.,

$$\boldsymbol{g} = \mathbf{F}\boldsymbol{x}. \tag{2.75}$$

Then, the covariance matrix of the parameters results as

$$\mathbf{C}_{gg} = \mathbf{F} \cdot \mathbf{C}_{xx} \cdot \mathbf{F}^{\text{T}}. \tag{2.76}$$

The distance between the vertices involved in a subgraph (e. g., a corner) is mostly quite small. Therefore, it seems reasonable to consider the distance-dependent correlations between these vertices. In that case, the secondary diagonal elements of $\mathbf{C}_{xx}$ are not zero. The hypothetic covariances can be calculated from (2.62) and (2.63).

In general, a statistical test consists of the following steps.

1. Stating a null hypothesis
2. Stating an alternative hypothesis
3. Calculating a test value
4. Calculating a critical value
5. Making the test decision.

In our case, the null hypothesis is the identity in the comparison

$$H_0\colon\ \boldsymbol{g}_1 = \boldsymbol{g}_2. \tag{2.77}$$

The alternative hypothesis is the disparity of the parameter vectors, i. e.,

$$H_{\text{alt}}\colon\ \boldsymbol{g}_1 \neq \boldsymbol{g}_2. \tag{2.78}$$

The test value is always a random value with a known distribution function. From the null hypothesis, we can conclude that the difference vector of both parameter vectors is zero, i. e.,

$$\boldsymbol{d} = \boldsymbol{g}_1 - \boldsymbol{g}_2 = 0. \tag{2.79}$$

The elements of the difference vector $\boldsymbol{d}$ are correlated random values. Its covariance matrix $\mathbf{C}_{dd}$ is calculated by variance propagation as

$$\mathbf{C}_{\text{dd}} = \mathbf{C}_{gg_1} + \mathbf{C}_{gg_2}. \tag{2.80}$$

With this covariance matrix, we are able to calculate a quadratic form with known χ^2 distribution as

$$\chi^2 = \boldsymbol{d}^{\mathrm{T}}\mathbf{C}_{\text{dd}}^{-1}\boldsymbol{d}. \tag{2.81}$$

If the subgraph is overparameterized—for instance, the case with directions expressed as normalized vectors—then the covariance matrix of the parameters is singular. In that case, an infinite number of inverse matrices exist. This problem can be solved if the pseudoinverse $\mathbf{C}_{dd}^{+}$ of $\mathbf{C}_{dd}$ is used, which is exactly the inverse with minimal trace. For this purpose, $\mathbf{C}_{dd}$ might be rendered with its eigenvectors of eigenvalue zero. In the case of a normalized direction vector, the eigenvector is identical to the direction vector itself.

The value χ^2 is χ^2 distributed, whereby the number of degrees of freedom is identical to the rank of $\mathbf{C}_{dd}$. In the case of a corner parameterized with a coordinate tuple and two direction vectors, the number of parameters is six, but the rank of $\mathbf{C}_{dd}$ is just four. If the same corner is parameterized by three coordinate tuples, then $\mathbf{C}_{dd}$ has the full rank of six.

The critical value χ^2_{c} is a function of the number of degrees of freedom and the significance level α. Common α values are 1 and 5%. The comparison of χ^2 with χ^2_{c} leads to the test decision.

If $\chi^2 < \chi^2_{\text{c}}$, then we accept the null hypothesis, therefore $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ are identical.

If $\chi^2 > \chi^2_{\text{c}}$, then we reject the null hypothesis in favor of the alternative hypothesis, and, therefore, $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ are not identical.

After testing all matching candidates, in principle, three results are possible.

1. No candidate was accepted as identical.
2. Exactly one candidate was accepted as identical.
3. More than one candidate was accepted as identical.

Solution 3 is ambiguous. In that case, all candidates should be rejected.

2.2.19 Search for Geometrical Constraints

Frequently, it is advisable to keep geometrical constraints during the adjustment calculation. It was shown in Sect. 2.2 how these constraints can be modeled by observations. However, before constraints can be introduced in an adjustment calculation, one has to know where they occur. In the same way that subgraphs of different datasets can be used to find related subgraphs, they can also be used within one dataset to find geometrical constraints. This is shown here with the example of rectangularity and collinearity constraints.

If a corner subgraph is parameterized by three coordinate tuples, then it can easily be tested whether the related vectors of coordinate differences are either rectangular or collinear. In the case of rectangularity, its scalar product has to be zero, i. e.,

$$\boldsymbol{v}_{\text{AB}} \cdot \boldsymbol{v}_{\text{BC}} \mathrel{\widehat{=}} \text{sp} \overset{!}{=} 0. \tag{2.82}$$

The scalar product sp is a function of the related coordinates as

$$\begin{aligned}\text{sp} &= f(\boldsymbol{g})\\ &= (x_{\text{A}} - x_{\text{B}})(x_{\text{C}} - x_{\text{B}}) + (y_{\text{A}} - y_{\text{B}})(y_{\text{C}} - y_{\text{B}}).\end{aligned} \tag{2.83}$$

It is a normally distributed random value whose standard deviation can be calculated by variance propagation as

$$\sigma_{\text{sp}}^2 = \mathbf{F} \cdot \mathbf{C}_{gg} \cdot \mathbf{F}^{\mathrm{T}} \quad \text{with} \quad \mathbf{F} = \frac{\text{dsp}}{\text{d}\boldsymbol{g}}. \tag{2.84}$$

Note that the covariance matrix of the related coordinates $\mathbf{C}_{gg}$ has nonzero secondary diagonal elements if distance-dependent correlations were modeled. With sp and σ_{sp}, a test value u can be calculated as

$$u = \frac{\text{sp}}{\sigma_{\text{sp}}}, \tag{2.85}$$

where u is normalized and normally distributed ($u \approx N(0,1)$). The null hypothesis is that u is equal to zero, i. e.,

$$H_0\colon\ u = 0. \tag{2.86}$$

The alternative hypothesis is that u is not equal to zero, which means that the test is a two-sided problem, i. e.,

$$H_{\text{alt}}\colon\ u \neq 0. \tag{2.87}$$

The critical value u_c is a function of the significance level α. It is calculated as the inverse of the distribution function of the normalized normal distribution as

$$u_c = \pm\Phi^{-1}\left(\frac{\alpha}{2}\right). \tag{2.88}$$

Common α values are 1 or 5%. Comparison of u with u_c leads to the test decision.

If $u < u_c$, then we accept the null hypothesis, and, therefore, u is zero, and the vectors are rectangular.

If $u > u_c$, then we reject the null hypothesis in favor of the alternative hypothesis, and, therefore, u is not zero, and the vectors are not rectangular.

In the case of collinearity, the length of the cross product of the vectors has to be zero, i. e.,

$$|\boldsymbol{v}_{AB} \times \boldsymbol{v}_{BC}| \mathrel{\widehat{=}} \text{cp} \overset{!}{=} 0. \tag{2.89}$$

The length of the cross product cp is a function of the related coordinates through

$$\begin{aligned}\text{cp} &= f(\boldsymbol{g})\\ &= (x_A - x_B)(y_C - y_B) - (y_A - y_B)(x_C - x_B).\end{aligned} \tag{2.90}$$

Like sp, also cp is normal distributed and can be tested for significance in the same way.

2.2.20 Interaction of Matching and Adjustment

Frequently, matching and adjustment are seen as completely independent processes. However, this view does not reflect reality. In fact, both processes interact. Datum-independent matching provides control points for the transformation. The transformation itself is an adjustment problem. After georeferencing, datum-dependent matching and a search for geometrical constraints can be accomplished. Results are additional identity constraint observations. With these observations, the observation vector of the adjustment can be extended, and a further adjustment calculation can be performed. The adjustment result gives information about mismatches. In an iterative process, false identity observations are removed. With the improved coordinates a new matching step is possible, and so on. Figure 2.25 shows the

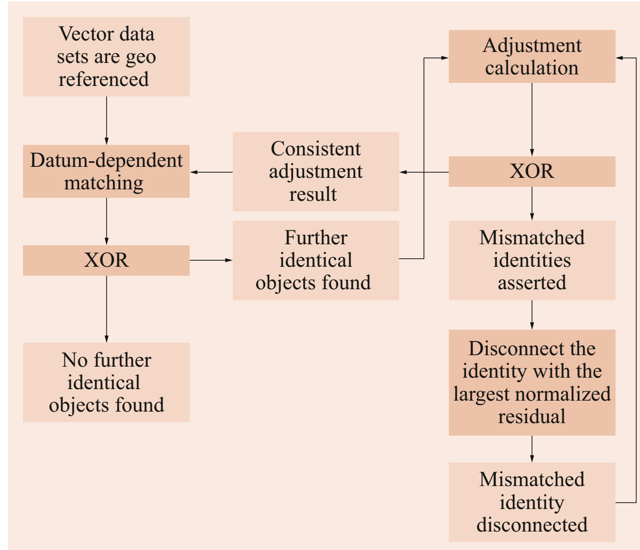


Fig. 2.25 Matching and adjustment in an event-driven process chain

interaction of matching and adjustment as an event-driven process chain [6]. As one can see, the interaction of matching and adjustment is an alternating iterative process whereby the adjustment itself constitutes an inner iteration loop.

2.3 Geostatistics

2.3.1 Geostatistical Analysis and Modeling

Originally applied to estimate mineral resources and reserves [7–9], geostatistics are nowadays used in various fields of research, including a wide range of applications in marine environmental sciences, like the spatial analysis and modeling of sedimentological features [10–12], benthic organisms [13–16], and fish stocks [17–24]. The theoretical foundation of geostatistical methods was developed by G. Matheron [8] in terms of the theory of regionalized variables. The theory assumes that observed values at defined locations in space are the realization of a random function, where each location is a random variable with an expected mean and variance. The observed values are, thereby, seen as residuals of the mean that may show spatial autocorrelation patterns across the area of interest. In geostatistics, the investigation and modeling of this spatial autocorrelation is done by variogram analytical techniques. The derived variogram model is then used as an input in the spatial interpolation process referred to as kriging.

Whether or not the expected mean of the random function can be seen as constant at each of the locations, different approaches have been developed. If the mean value is assumed to be known and constant over all locations, simple kriging may be applied, if it is unknown and constant, ordi-

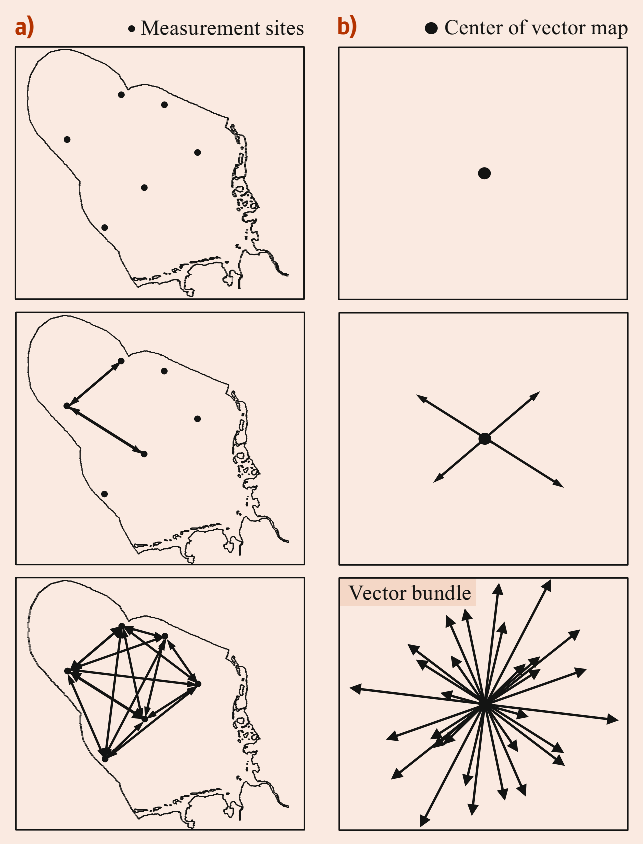


Fig. 2.26 Forming distance vectors from point pairs: **a** study area, **b** virtual vector map

nary kriging may be used. Provided that the expected mean is assumed to vary between sites, this so called drift can be modeled in different ways. One possibility to calculate such a trend surface is to apply polynomial functions using the spatial coordinates of the observation sites as explanatory variables [25, 26]. Measured values are subtracted from this trend surface and the resulting residuals investigated for spatial autocorrelation via variogram analysis and interpolated via kriging. The resulting residual map is then combined with the trend surface map to obtain the spatial estimation result. If the coordinates of the observations are used to calculate the trend surface, the geostatistical method is commonly referred to as universal kriging [25]. Another possibility to model a trend surface is to use full coverage auxiliary information on variables that are sufficiently correlated to the observed values. This geostatistical approach is referred to as regression kriging [27–31], where the modeling of the trend surface is done by applying different statistical methods like, e. g., linear regression analysis. The authors are aware that both approaches are sometimes summarized under universal kriging but prefer to differentiate into universal and regression kriging within the scope of this book chapter. A variety of literature mentioning different kriging terms and its slight distinctions is summarized by [28].

Semivariogram maps or surfaces can be used to calculate an experimental semivariogram from a given dataset [25, 32, 33]. Variogram maps allow us to account for so-called anisotropies in the data field. Anisotropies are properties of the underlying spatial process where spatial dependence not only changes with distance but also with direction. Examples of such processes can be found especially in marine environmental applications, where ecological patterns strongly depend upon topographical or hydrographical conditions. The following work steps can be distinguished in modeling the spatial autocorrelation structure with the aid of variogram maps [25, pp. 253–255]:

- Forming direction vectors also referred to as lag vectors from all sample point pairs of the considered sample point distribution (Fig. 2.26a,b).
- Connection of the origins of all lag vectors in a virtual vector map resulting in a variogram bundle (Fig. 2.26c). Note that variogram maps are point-symmetrical relative to their origin because all lag vectors have a counterpart in the opposite direction.
- Overlaying the vector map with a grid of defined mesh size (Fig. 2.27a) and assigning semivariances for each raster cell or bin (Fig. 2.27b):

$$\gamma(h) = \frac{1}{2N h)} \sum_{i=1}^{N} (z_i - z_{i+h})^2 \qquad (2.91)$$

  where $\gamma(h)$ is the semivariance for a certain bin, $N(h)$ the number of sample pairs (or vectors) for a certain bin, and z_i, z_{i+h} pairs of measured values of each vector for a certain bin.
  The semivariances are calculated over all sample pairs that fall within and near a respective bin. Each squared difference is thereby weighted according to a Kernel function depending on the distance of the ending of each lag vector to the center of each bin. This way, a semivariogram map or surface is derived from the measurement values (Fig. 2.27b). For a more detailed description of the underlying methodology please refer to [19, pp. 253–255].
- Assigning the calculated semivariances to a coordinate system defined by the distance of the origin of the variogram map to the center of each bin (abscissa) and the semivariances that were calculated for the respective bins of the variogram map (ordinate). This way, an experimental variogram is calculated (Fig. 2.27c).
- In order to perform kriging it is necessary to fit a defined variogram model to the experimental variogram. This can be achieved by means of mathematical models fitted to the experimental variogram in terms of a least-squares regression line (Fig. 2.27c). Commonly used model types are, e. g., spherical and exponential variogram models.

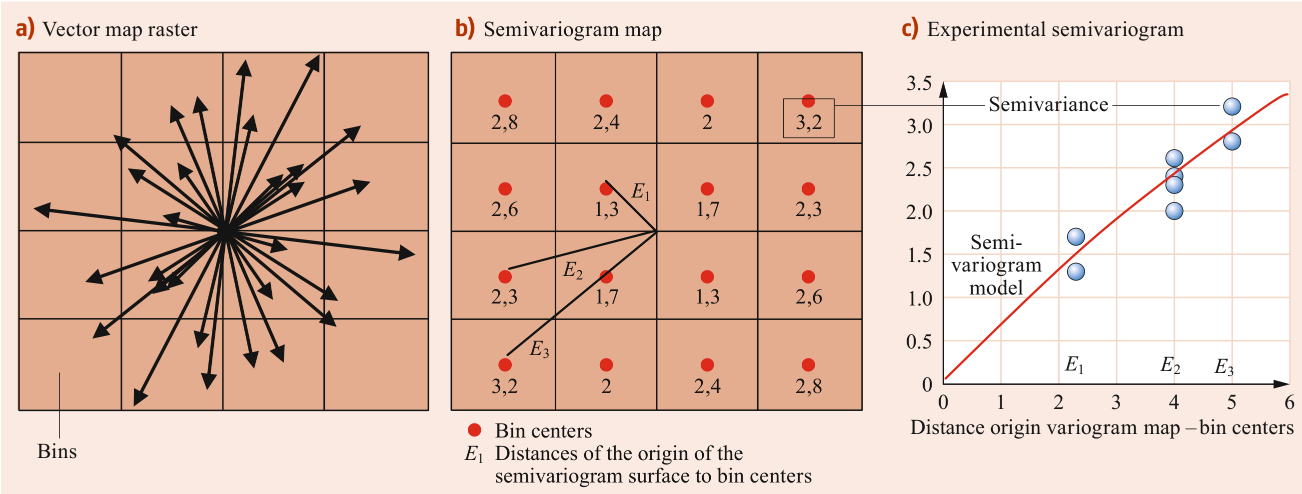


Fig. 2.27 From the vector map to the experimental semivariogram

Variogram models are usually described with respect to three significant measures: the nugget effect, sill and range (Fig. 2.28). The range equals the maximum separation distance within which a distinct increase of semivariogram values can be observed. Within this distance, spatial autocorrelation is indicated. The sill corresponds to the semivariance assigned to the range. If anisotropies can be detected, both sill and range may vary with respect to direction. Local variabilities of the observed variable or measurement errors lead to high semivariances at nearby locations being reflected in terms of the so-called nugget effect in the variogram model, where the variogram model cuts the ordinate above the origin. According to [34] the nugget/sill ratio is a significant measure for evaluating the strength of autocorrelation and, thus, the appropriateness of using the variogram for geostatistical estimation. Whether or not anisotropies are detected determines whether this kriging window is circular (no anisotropies) or ellipsoidal (different ranges of spatial autocorrelation for different directions). In the latter case, different variogram models with different ranges are calculated for different directions.

Cross validation should be used for choosing an adequate variogram model and for describing the quality of estimation. For this purpose, each measurement value is extracted from the dataset separately, and the selected variogram model is used to estimate the measurement variable at the exact same location via kriging. By subtracting each measured value from each corresponding estimated value an error can then be calculated, resulting in an error distribution for the whole dataset. Various key parameters can be calculated from this distribution to characterize the global quality of the chosen variogram model and the estimation as a whole. Examples of such parameters calculated for this study are the mean error (ME – the average value of the cross-validation errors, which at best should be 0) or the absolute error (AE – the mean of all absolute values of the cross-validation errors, which should be as small as possible).

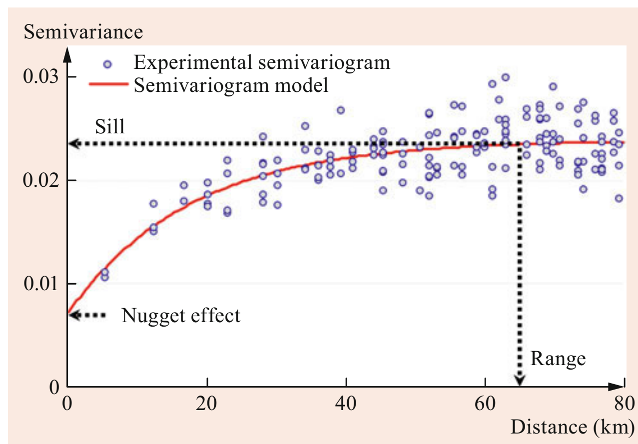


Fig. 2.28 Example of a semivariogram model

2.3.2 Background

To account for future challenges of marine spatial planning and nature conservation in Europe full coverage data on specific features of the benthic environment are needed. These needs arise from European Union directives, e. g., the Marine Strategy Framework Directive (Council Directive 2008/56/EC [35], Commission Decision 2017/848 [36]) or the Habitats Directive 92/43/EEC [37], and national legislations like the Federal German Nature Conservation Law (BNatSchG, Bundesnaturschutzgesetz). In Germany's Exclusive Economic Zone (EEZ) of the North Sea, benthic biotope maps are currently being developed in a national project funded by the Federal Agency for Nature Conservation [38, 39]. The goal is to map biotopes according to the classification system of the Red List of Threatened Biotopes by [40] and protected biotopes of § 30 of the Federal German Nature Conservation Law comprising reefs, sandbanks, muddy bottoms with burrowing megafauna and so-called Species-rich gravel, coarse sand, and shell gravel areas in the EEZ of the North Sea. The mapping is done with respect to different spatial resolutions ranging from high-resolution maps (100 m × 100 m) to broad scale maps (1000 m × 1000 m) and mainly relies on punctual biological sample data on infauna and epibenthic species, as well as full-coverage abiotic data on sediments, topography, and hydrography. Data on the infauna (animals that mainly live in the sediment) are sampled with help of grabs (van Veen grabs); data on epibenthos (animals and plants that live on the sea floor) are acquired by the use of, e. g., towed dredges and video sledges.

Different modeling options exist to derive full-coverage raster maps from sample point data, including the application of geostatistical techniques and predictive modeling approaches. In a biological context, the spatial prediction of a species or community by use of biological sample data and full-coverage environmental data is commonly referred to as species distribution or ecological niche modeling. A wide range of corresponding literature has been published on related theoretical concepts and methodologies, which will not be discussed in this book chapter. For comprehensive reviews on the subject, please refer to, e. g., [30, 41–46]. Within the German EEZ and adjacent coastal areas, species distribution models have been applied to predict and map the distribution of benthic communities [39, 47–49].

In a recently published case study, predictive modeling techniques were used to estimate the spatial distribution of the above-mentioned § 30 biotope Species-rich gravel, coarse sand, and shell gravel areas [50]. For this, point data on the occurrences of certain characteristic species and sedimentological variables were intersected with full-coverage sedimentological and topographical data. Random Forests [51] was then applied to map potential spatial patterns of the protected biotope within the two Natura 2000 sites Sylt Outer Reef and Borkum Reef Ground. From a biological perspective, the results proved to be plausible and model performance and validation criteria showed good results. Nonetheless, the spatial patterns should be substantiated by the use of additional biological and environmental data and by applying alternative modeling techniques to Random Forests.

The book chapter at hand partly contributes to this goal by applying geostatistical methods to map the benthic soft-bottom community Goniadella-Spisula within the EEZ of the North Sea and adjacent coastal areas. The respective community mostly occurs on coarse sediments and is indicative for the incidence of the § 30 biotope Species-rich gravel, coarse sand, and shell gravel areas in the North Sea [52]. The study at hand introduces two geostatistical methods to predict the possible occurrence of the Goniadella-Spisula community within the area of interest: universal kriging and regression kriging. Both methods account for the spatial autocorrelation of the observed values in the interpolation process and include full-coverage information to access the so-called drift in terms of the spatially varying expected value of an underlying random function.

2.3.3 Data and Procedure

The data used to map the spatial distribution of the Goniadella-Spisula community were taken from the initial phase of the national biotope mapping project between 2011 and 2014, financed and coordinated by the Federal Agency for Nature Conservation [38, 39]. Within this first phase of the project, six benthic soft-bottom communities were derived from abundance data on infauna organisms sampled at 1146 benthic monitoring stations via van Veen grabs. The corresponding sampling was performed both within the biotope mapping project itself and other national biotope and monitoring activities between 2000 and 2014. As commonly applied in benthic biology, cluster analysis was used to derive benthic communities from the abundance data. Each resulting cluster was thereby analyzed in terms of typical or characteristic species, which were then compared to species of commonly known benthic soft-bottom communities within the German EEZ and coastal areas of the North Sea. Here, a fuzzy clustering approach was decided on to calculate the benthic soft-bottom communities. As a result, the uncertainty of assigning a given cluster to a chosen monitoring station could be provided, ranging between 0 (minimum strength of memberships to a given cluster) to 1 (maximum strength of memberships to a given cluster). Within the scope of this chapter, the fuzzy membership scores of the soft bottom community Goniadella-Spisula were used as the variable of interest in the geostatistical modeling procedure. Please refer to Fig. 2.29 for an overview on the spatial distribution of the respective fuzzy membership scores in the German EEZ and coastal areas. For a comprehensive explanation on the application of fuzzy clustering in this specific context, please, furthermore, refer to [53].

The geostatistical calculations, including variogram analyses as well as kriging and cross-validation procedures, were performed by the Geostatistical Analyst extension from Esri ArcGIS 10.2.2 [25] using the Universal Transverse Mercator (UTM) coordinate system (Zone 32) as a spatial reference system. The universal kriging approach could be completely performed by tools available in the ArcGIS Geostatistical Analyst extension (providing different polynomial functions available to calculate the trend surface). Regression kriging was carried out both in ArcGIS and in the open-source

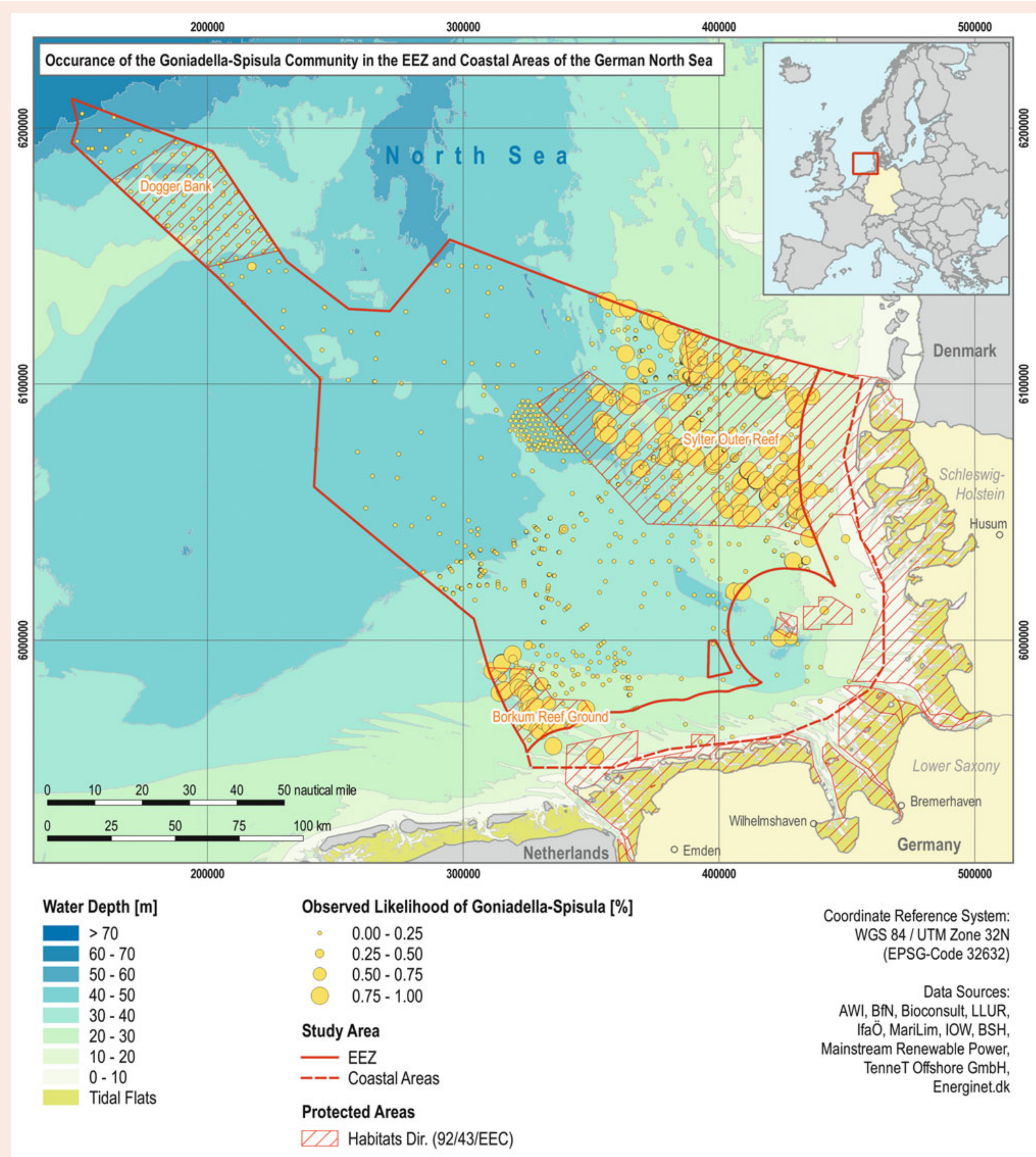


Fig. 2.29 Fuzzy membership scores for the soft-bottom community Goniadella-Spisula within the German EEZ and coastal areas of the North Sea [38]

software R [54], where linear regression analysis was applied to derive a regression model for the fuzzy membership values of Goniadella-Spisula by auxiliary full-coverage information on variables that are assumed to have an impact on the occurrence of this community. For this, respective monitoring stations were intersected with full-coverage data on coarse sediments derived from [55], interpolated grain-size fractions for sand, muddy, and coarse sediments [56], bathymetry, slope, and information on the delineation of reefs [57]. The bathymetry data layer was provided by

EMODnet Seabed Habitats [58], slope was derived by [56] from the bathymetry layer for different radiuses using the quadratic surface approximation by [59] as implemented in the open-source software SAGA-GIS [60]. All full-coverage data were either resampled or rasterized (from polygon data) to a 230 m × 230 m raster referring to the highest spatial resolution, which was provided by the raster data on bathymetry. Linear regression analysis was then applied to derive a regression model for the Goniadella-Spisula membership values within the statistical open-source software R v. 3.4.0 using the stats package [54]. The final regression model only included variables of a high level of significance ($p < 0.001$) and was applied on the full-coverage data to compute the trend surface map. Variogram analysis was applied to examine and model the spatial autocorrelation of the residuals and to calculate a residual map via kriging. As a final result, residual and trend surface map were combined to the overall regression kriging map for the incidences of the Goniadella-Spisula community within the German North Sea area.

For both the universal and the regression kriging approach, the mean distance of each measuring site to its nearest neighbor was chosen as the initial value to determine an adequate lag size for the calculation of a variogram model [25]. To optimally model the autocorrelation structures a multitude of isotropic and anisotropic variogram models (including different ranges, sills, nugget effects) were then compared with regard to the results of the cross validation. According to [25] it was thereby aimed to visualize the variogram model, so that the increase of semivariances could be clearly identified in the variogram window over ca. two-thirds of the distance axis. A four-sector neighborhood was defined as a kriging window to avoid directional bias, and a maximum of ten points were accounted for in each sector to estimate a value at a certain point.

2.3.4 Results and Outlook

The results of the universal kriging approach, including semivariogram, variogram map, cross-validation measures and mapping result are illustrated in Fig. 2.30. As a trend surface a second-order polynomial function was decided on and an anisotropic spherical variogram model was chosen to best model the spatial autocorrelation structure. The corresponding model shows a low nugget/sill ratio of 17%; the mean of the cross-validation errors is 0.01 and the mean absolute error is 0.11. Anisotropies could be detected in the variogram map, so that different ranges were calculated for different directions ranging from 37.5 km in north–west direction to 18 km in north–east direction. The resulting universal kriging map clearly shows the effect of these anisotropies in the estimated probability of occurrences of the Goniadella-Spisula community. These show highest values within and near the Natura 2000 sites Sylt Outer Reef and Borkum Reef Ground, where Pleistocene coarse sediments and stone fields cover the sea floor.

The final regression kriging map, including semivariogram, variogram map, and cross-validation measures of the residuals is illustrated in Fig. 2.31. As can be seen by the results of variogram analysis, the spatial dependence between residuals only allowed the modeling of an isotropic spatial autocorrelation structure with a nugget/sill ratio of 64%, which, however, still indicates spatial autocorrelation of medium strength according to [61]. The mean of the cross-validation errors is less than 10^{-4} and the mean absolute error is 0.135. The derived regression model used for the trend surface map revealed an R^2 of 0.45, meaning that 45% of variance of the fuzzy membership values of Goniadella-Spisula could be explained by the auxiliary variables. As the only significant variables ($p < 0.001$) bathymetry, interpolated coarse sediments, as well as the incidences of reefs and coarse sediments provided by [57] and [55], respectively, were included. To evaluate the surface estimation of the final regression kriging map, a root-mean-square error (RMSE) between estimated and observed values of 0.2 was calculated according to [31]. An RMSE close to zero thereby indicates a good estimation quality, meaning that average estimated and observed values are very similar. As an additional validation procedure, the linear relationship between observed and estimated values was examined by regression analysis. The resulting model showed an R^2 of 0.64. This again indicates a good correspondence between the original Goniadella-Spisula fuzzy membership values and the results of spatial estimation.

Comparing the two predicted surface maps, differences can mainly be found concerning the spatial structure of areas of the probability of the occurrence of communities above 0.5. Whereas the universal kriging approach led to rather contiguous areas, the regression kriging approach predicted rather mosaic-like patterns. The latter thereby clearly reflects the spatial structure of the occurrence of coarse sediments as derived by [55] which showed a high impact in the results of regression analysis performed to derive the trend surface map. In general, areas with predicted membership values of above 0.5 do not vary significantly between both approaches (universal kriging: 1745 km^2, regression kriging: 1889 km^2). Overall, both kriging approaches led to similar and plausible spatial predictions.

In this section, two alternative geostatistical approaches were applied to map the occurrences of the Goniadella-Spisula community within the German EEZ and adjacent coastal areas. As this community is strongly associated to the § 30 biotope Species-rich gravel, coarse sand, and shell gravel areas, the mapping results can be seen as indicative for the occurrence of this protected biotope in the German

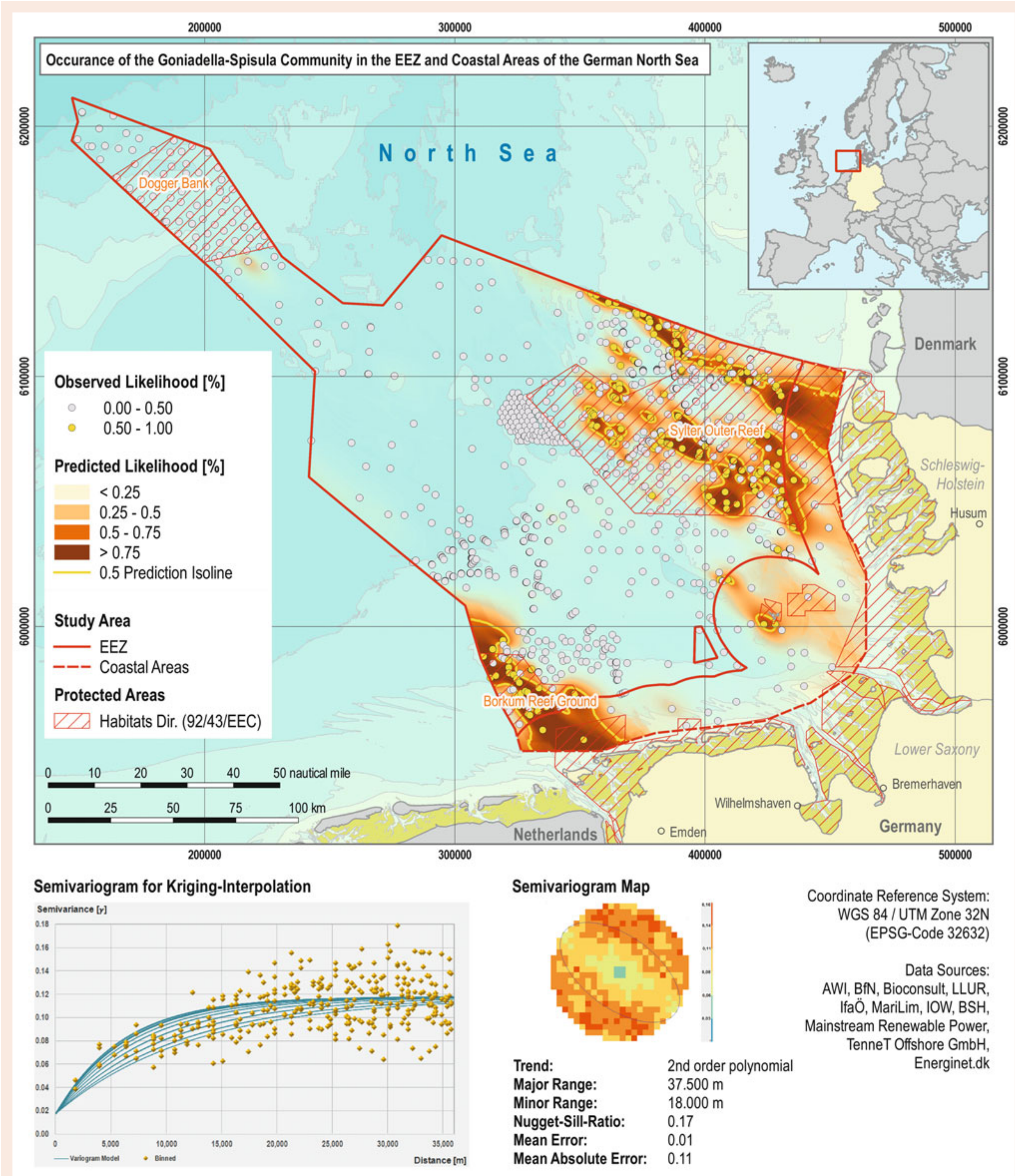


Fig. 2.30 Spatial estimation of the probability of occurrence of the Goniadella-Spisula community by universal kriging, including semivariogram, semivariogram map, and cross-validation measures

North Sea area. Since the different predictive modeling results of this biotope presented by [50] show very similar spatial patterns, the results at hand can be used to substantiate the corresponding findings. In a second phase of the German biotope mapping project between 2015 and 2018, additional infauna abundance data and full coverage data

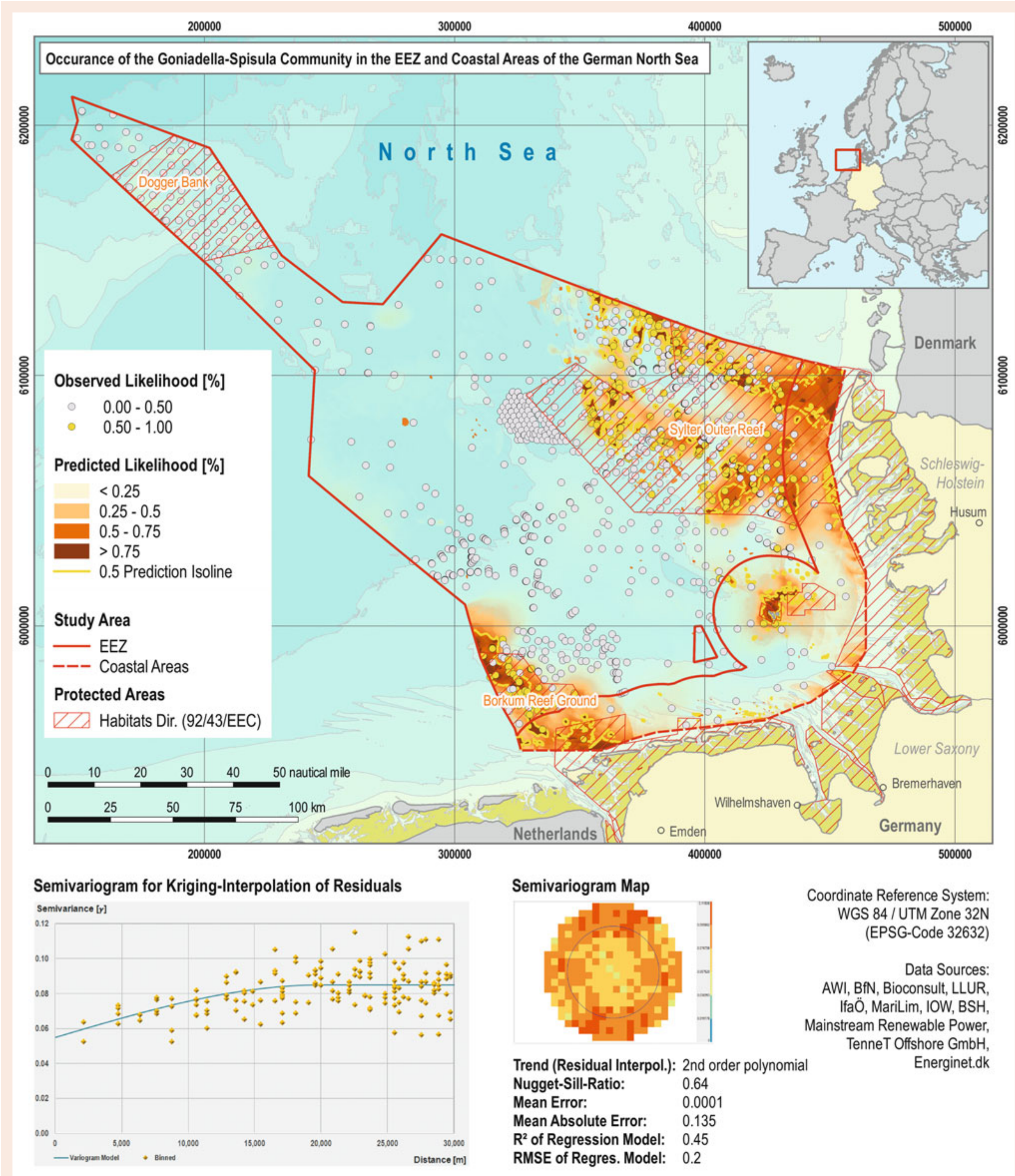


Fig. 2.31 Spatial estimation of the probability of occurrence for the Goniadella-Spisula community by regression kriging, including semivariogram, semivariogram map, and cross-validation measures

on auxiliary information was produced. As one important full-coverage data set high-resolution sediment data acquired from hydroacoustic investigations will be available. Hence, the geostatistical modeling of this specific soft-bottom community could be repeated in order to evaluate and update the maps presented in this book chapter.

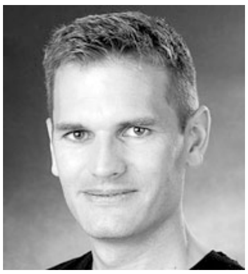
Acknowledgements The works presented in this study were produced within a national biotope mapping project coordinated and financed by the German Federal Agency for Nature Conservation (Bundesamt für Naturschutz, BfN). Several other institutions provided measurement data on macrozoobenthos and sediments as well as full coverage data on sediments and bathymetry. These include BioConsult Schuchardt & Scholle GbR, the Alfred-Wegener-Institute for Polar and Marine Research (Bremerhaven), the Federal Maritime Hydrographic Office (Bundesamt für Seeschiffahrt und Hydrographie BSH), Leibniz Institute for Baltic Sea Research (Institut für Ostseeforschung, Warnemünde) and the federal states environmental authorities Landesamt für Landwirtschaft, Umwelt und ländliche Räume Schleswig-Holstein (LLUR) and Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz (NLWKN).

References

1. Mikhail, E., Ackermann, F.: Observations and Least Squares. Univ. Press America, New York (1976)
2. Niemeier, W.: Ausgleichungsrechnung. De Gruyter, Berlin (2002)
3. Conformal map. http://en.wikipedia.org/wiki/Conformal_map
4. Student's t-test. http://en.wikipedia.org/wiki/T-test
5. . http://en.wikipedia.org/wiki/Graph_(mathematics)
6. Event-driven process chain. http://en.wikipedia.org/wiki/Event-driven_process_chain
7. Krige, D.G.: A statistical approach to some basic mine valuation problems on the Witwatersrand. J. S. Afr. Inst. Min. Metall. **52**(6), 119–139 (1951)
8. Matheron, G.: Les variables régionalisées et leur estimation. Masson, Paris (1965)
9. Matheron, G.: The Theory of Regionalized Variables and Its Applications. Les Cahiers du Centre de Morphologie Mathematique in Fontainebleu, Paris (1971)
10. Diesing, M., Green, S.L., Stephens, D., Lark, R.M., Stewart, H.A., Dove, D.: Mapping seabed sediments: comparison of manual, geostatistical, object-based image analysis and machine learning approaches. Cont Shelf Res **84**, 107–119 (2014)
11. Jerosch, K.: Geostatistical mapping and spatial variability of surficial sediment types on the Beaufort Shelf based on grain size data. J. Mar. Syst. **127**, 5–13 (2013)
12. Meilianda, E., Huhn, K., Alfian, D., Bartholomae, A.: Application of multivariate geostatistics to investigate the surface sediment distribution of the high-energy and shallow Sandy Spiekeroog shelf at the German bight, southern North Sea. Open J. Mar. Sci. **2**(04), 103 (2012)
13. Guarini, J.M., Blanchard, G.F., Bacher, C., Gros, P., Riera, P., Richard, P., Gouleau, D., Galois, R., Prou, J., Sauriau, P.G.: Dynamics of spatial patterns of microphytobenthic biomass: inferences from a geostatistical analysis of two comprehensive surveys in Marennes-Oléron Bay (France). Mar. Ecol. Prog. Ser. **166**, 131–141 (1998)
14. Maynou, F.: The application of geostatistics in mapping and assessment of demersal resources, Nephropsnorvegicus in the northwestern Mediterranean: a case study. Sci. Mar. **62**(1), 117–133 (1998)
15. Sørensen, P.S.: Spatial Distribution Maps for Benthic Communities. Institute of Mathematical Modelling, Technical Univ. of Denmark, Lyngby (2000)
16. Souza, F.M.D., Gilbert, E.R., Camargo, M.G.D., Pieper, W.W.: The spatial distribution of the subtidal benthic macrofauna and its relationship with environmental factors using geostatistical tools: a case study in Trapandé Bay, southern Brazil. Zoologia **30**(1), 55–65 (2013)
17. Aidoo, E.N., Mueller, U., Goovaerts, P., Hyndes, G.A.: Evaluation of geostatistical estimators and their applicability to characterise the spatial patterns of recreational fishing catch rates. Fish. Res. **168**, 20–32 (2015)
18. Petitgas, P.: Geostatistics for fish stock assessments: a review and an acoustic application. ICES J. Mar. Sci. **50**(3), 285–298 (1993)
19. Petitgas, P.: Geostatistics in fisheries survey design and stock assessment: models, variances and applications. Fish Fish. **2**(3), 231–249 (2001)
20. Petitgas, P.: Geostatistics and their applications to fisheries survey data: a history of ideas, 1990–2007. In: Computers in Fisheries Research, pp. 191–224. (2009)
21. Petitgas, P., Woillez, M., Rivoirard, J., Renard, D., Bez, N.: Handbook of Geo-statistics in R for Fisheries and Marine Ecology. ICES cooperative research report, vol. 338. (2017)
22. Rufino, M.M., Stelzenmüller, V., Maynou, F., Zauke, G.P.: Assessing the performance of linear geostatistical tools applied to artificial fisheries data. Fish. Res. **82**(1–3), 263–279 (2006)
23. Stelzenmüller, V., Maynou, F., Bernard, G., Cadiou, G., Camilleri, M., Crec'hriou, R., Criquet, G., Dimech, M., Esparzad, O., Higgins, R., Pérez-Ruzafad, A., Lenfant, P.: Spatial assessment of fishing effort around European marine reserves: implications for successful fisheries management. Mar. Pollut. Bull. **56**(12), 2018–2026 (2008)
24. Vaz, S., Martin, C.S., Ernande, B., Coppin, F., Harrop, S., Carpentier, A.: Using geostatistics to quantify annual distribution and aggregation patterns of fishes in the eastern English Channel. In: Proc ICES AnnuSciConf (Aberdeen), pp. 20–24. (2005)
25. Johnston, K., VerHoef, J.M., Krivoruchko, K., Lucas, N.: Using ArcGIS Geostatistical Analyst. ESRI, Redlands (2001)
26. Wackernagel, H.: Multivariate Geostatistics: An introduction with Applications. Springer, Berlin, Heidelberg, New York (2003)
27. Hengl, T.: A Practical Guide to Geostatistical Mapping of Environmental Variables. EUR 22904 EN-scientific and technical research series. Office for Official Publications of the European Communities, Luxemburg (2007)
28. Hengl, T., Heuvelink, G.B., Stein, A.: Comparison of Kriging with External Drift and Regression Kriging. ITC, Enschede (2003)
29. Hengl, T., Heuvelink, G.B., Stein, A.: A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma **120**(1–2), 75–93 (2004)
30. Hengl, T., Sierdsema, H., Radović, A., Dilo, A.: Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. Ecol. Model. **220**(24), 3499–3511 (2009)
31. Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J.: Further results on prediction of soil properties from terrain attributes: heterotropic cokriging and regression-kriging. Geoderma **67**, 215–226 (1995)
32. Gribov, A., Krivoruchko, K., VerHoef, J.M.: Modeling the semivariogram: New approach, methods comparison, and simulation study. In: Coburn, T.C., Yarus, J.M., Chambers, R.L. (eds.) Stochastic Modeling and Geostatistics: Principles Methods, and Case Studies, Volume II AAPGcomputer applications in geology, vol. 5, pp. 45–57. (2006)
33. Lawrence, J., Southworth, S., Sutphin, D.M., Rubis, G.A., Schuenemeyer, J.H., Burton, W.C.: Validation of the relation between structural patterns in fractured bedrock and structural information interpreted from 2D-Variogram maps of water-well yields in Loudoun county, Virginia. Nat. Resour. Res. **13**(4), 255–264 (2004)
34. Cambardella, C.A., Moorman, T.B., Parkin, T.B., Karlen, D.L., Novak, J.M., Turco, R.F., Konopka, A.E.: Field-scale variability of soil properties in central Iowa soils. Soil Sci. Soc. Am. J. **58**(5), 1501–1511 (1994)
35. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action

in the field of marine environmental policy (Marine Strategy Framework Directive). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32008L0056. Official Journal of the European Union L 164/19
36. Commission Decision (EU) 2017/848 of 17 May 2017 laying down criteria and methodological standards on good environmental status of marine waters and specifications and standardised methods for monitoring and assessment, and repealing Decision 2010/477/EU. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1495097018132&uri=CELEX:32017D0848. Official Journal of the European Union L125/43
37. Council Directive 92/43/EEC of 21 May 1992on the conservation of natural habitats and of wild fauna and flora
38. Bildstein, T., Fiorentino, D., Günther, C.P., Pesch, R., Rückert, P., Schröder, W., Schuchardt, B.: Cluster 6 Biotopkartierung: Endberichtsentwurf – Teil Nordsee (2014). unveröff. Bericht i. A. des Bundesamtes für Naturschutz (BfN)
39. Pesch, R., Propp, C., Darr, A., Bartholomä, A., Beisiegel, K., Bildstein, T., Fiorentino, D., Hass, C., Holler, P., Lambers-Huesmann, M., Richter, P., Papenmeier, S., Günther, C.-P., Schiele, K., Schuchardt, B., Schwarzer, K., Tauber, F., Zeiler, M., Zettler, M.: Progress in marine biotope mapping in Germany. In: v. Nordheim, H., Wollny Goerke, K. (eds.) Progress in Marine Conservation in Europe 2015 BfN-Skripten, vol. 451, pp. 115–120. Bundesamt für Naturschutz, Bonn (2016)
40. Finck, P., Heinze, S., Raths, U., Riecken, U., Ssymank, A.: Rote Liste der gefährdeten Biotoptypen Deutschlands. Dritte fortgeschriebene Fassung 2017. Bundesamt für Naturschutz, Bonn (2017)
41. Austin, M.: Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecol. Model. **200**(1–2), 1–19 (2007)
42. Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B., Singer, A.: Correlation and process in species distribution models: bridging a dichotomy. J Biogeogr. **39**(12), 2119–2131 (2012)
43. Elith, J., Leathwick, J.R.: Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. **40**, 677–697 (2009)
44. Guisan, A., Thuiller, W.: Predicting species distribution: offering more than simple habitat models. Ecol Letters **8**(9), 993–1009 (2005)
45. Robinson, L.M., Elith, J., Hobday, A.J., Pearson, R.G., Kendall, B.E., Possingham, H.P., Richardson, A.J.: Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. Glob. Ecol. Biogeogr. **20**, 789–802 (2011)
46. Robinson, N.M., Nelson, W.A., Costello, M.J., Sutherland, J.E., Lundquist, J.C.: A systematic review of marine-based Species Distribution Models (SDMs) with recommendations for best practice. Front. Mar. Sci. **4**(421), 1–11 (2017). 10.3389/fmars.2017.00421
47. Neumann, H., Diekmann, R., Emeis, K.C., Kleeberg, U., Moll, A., Kröncke, I.: Full-coverage spatial distribution of epibenthic communities in the south-eastern North Sea in relation to habitat characteristics and fishing effort. Mar. Environ. Res. **130**, 1–11 (2017)
48. Pesch, R., Pehlke, H., Jerosch, K., Schröder, W., Schlüter, M.: Using decision trees to predict benthic communities within and near the German Exclusive Economic Zone (EEZ) of the North Sea. Environ Monit Assess **136**(1–3), 313–325 (2008)
49. Singer, A., Schückel, U., Beck, M., Bleich, O., Brumsack, H.J., Freund, H., Geimecke, C., Lettmann, K.A., Millat, G., Staneva, J., Vanselow, A., Westphal, H., Wolff, J.O., Wurpts, A., Kröncke, I.: Small-scale benthos distribution modelling in a North Sea tidal basin in response to climatic and environmental changes (1970s–2009). Mar. Ecol. Prog. Ser. **551**, 13–30 (2016)
50. BioConsul: Kleinmaßstäbige Abgrenzung des nach §30 BNatSchG geschützten Biotoptyps "Artenreiche Kies-, Grobsand- und Schillgründe" in den FFH-Gebieten der AWZ der Nordsee. Bundesamt für Naturschutz, Vilm (2017). Erstellt im Rahmen des FuE-Vorhabens des BfN: Erfassung, Bewertung und Kartierung benthischer Arten und Biotope (AWZ-P4, Benthos)
51. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
52. BfN: Artenreiche Kies-, Grobsand- und Schillgründe im Meeres- und Küstenbereich – Definition und Kartieranleitung Kies-, Grobsand- & Schillgründe (2011). http://www.bfn.de/habitatmare/de/downloads-marine-biotope.php, Accessed 30 May 2012
53. Fiorentino, D., Pesch, R., Guenther, C.P., Gutow, L., Holstein, J., Dannheim, J., Ebbe, B., Bildstein, T., Schroeder, W., Schuchardt, B., Brey, T., Wiltshire, K.H.: A 'fuzzy clustering' approach to conceptual confusion: how to classify natural ecological associations. Mar Ecol. Prog. Ser. **584**, 17–30 (2017). 10.3354/meps12354
54. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna (2015). https://www.R-project.org/
55. Laurer, W.-U., Naumann, M., Zeiler, M.: Sedimentverteilung auf dem Meeresboden in der deutschen Nordsee nach der Klassifikation von FIGGE (1981). Geopotential Deutsche Nordsee, LBEG – BSH – BGR, Hannover/Hamburg (2014)
56. S. Schönrock: Comparison of predictive statistical methods for full coverage mapping of benthic soft bottom communities within the Exclusive Economic Zone (EEZ) of the German North Sea, Masters Thesis (Beuth University of Applied Sciences, Berlin 2016), originally written in German
57. BfN: Karte 3: Verteilung der abgrenzungsrelevanten FFH-Schutzgüter sowie die FFH-Gebietsmeldungen "Doggerbank" (DE 1003-301); "Sylter Außenriff" (DE 1209-301); "Borkum Riffgrund" (DE 2104-301) in der AWZ der deutschen Nordsee; Stand 28.04.2004: 1 (2004)
58. Populus, J., Vasquez, M., Albrecht, J., Manca, E., Agnesi, S., Al Hamdani, Z., Andersen, J., Annunziatellis, A., Bekkby, T., Bruschi, A., Doncheva, V., Drakopoulou, V., Duncan, G., Inghilesi, R., Kyriakidou, C., Lalli, F., Lillis, H., Mo, G., Muresan, M., Salomidi, M., Sakellariou, D., Simboura, M., Teaca, A., Tezcan, D., Todorova, V., Tunesi, L.: EUSeaMap, A European broad-scale seabed habitat map (2017). 10.13155/49975
59. J. Wood (1996): The Geomorphological characterisation of Digital Elevation Models, Ph.D. thesis (Department of Geography, University of Leicester 1996)
60. Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J.: System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. **8**(7), 1991–2007 (2015)
61. Zhao, B., Cai, Q.: Geostatistical analysis of chlorophyll a in freshwater ecosystems. J. Freshw. Ecol. **19**(4), 613–621 (2004)
62. Schönrock, S., Schuchardt, B., Bildstein, T., Kreutle, A., Heinicke, K., Pesch, R.: Geostatistical Applications in a Marine Benthic Biological Context. – Kresse, W., Danko, D. M. (Eds.): Springer Handbook of Geographic Information. Springer (2020)

Frank Gielsdorf Dr Frank Gielsdorf graduated in 1987 as Dipl.-Ing. in Surveying from the Technical University of Dresden, Germany and obtained his Doctoral degree in 1997 from the Technical University of Berlin. From 1995 to 2006 he was Assistant Professor in the Department of Geodesy and Geoinformation at the Technical University of Berlin. Since May 2006 he has been working as a developer for technet GmbH Berlin, Germany.



Simon Schönrock Simon Schönrock holds a Bachelor's degree in Geopraphy and a Master's degree in Geoinformation and is a certified GIS Analyst. He investigated modeling approaches for spatial prediction of benthic habitats in the German North Sea. He currently works for the Institut biota GmbH in north-eastern Germany, with responsibilities in geospatial and statistical analysis, focusing on management, monitoring and evaluation of surface waters.

Roland Pesch Roland Pesch is Professor for Foundations and Applications of Geoinformation Systems at Jade University of Applied Sciences in Oldenburg. As an environmental geodata analyst he works on different monitoring, assessment and planing issues in terrestrial and marine environments. His main research thereby concerns the analysis and mapping of ecological phenomena on different spatial scales, applying spatial statistics and GIS technologies.

3 Databases

Thomas Brinkhoff and Wolfgang Kresse

Contents

Abstract

As geographic information storage and applications matured, their use as databases followed. A typical geoconfiguration consists of a map combined with an object-relational database, similar to the 300 year-old example shown in Fig. 3.1. Other geographic databases, such as the well-known Earth browsers Bing or Google Maps, contain a simple, but large, collection of raster orthophoto maps. Vector maps require a far more sophisticated data model and are usually rendered while being read from the database and presented on a display device.

Sections 3.1–3.5 provide basic knowledge about database theory. The two most common models, namely the relational and the object-oriented model, are explained. The second part of this chapter (Sects. 3.6–3.12) explains the geospecific aspects of database technol-

T. Brinkhoff (✉)
Institute for Applied Photogrammetry and Geoinformatics (IAPG), Jade University of Applied Sciences
Ofener Straße, 16/19, 26121 Oldenburg, Germany
e-mail: thomas.brinkhoff@jade-hs.de

W. Kresse
University of Applied Sciences Neubrandenburg
Neubrandenburg, Germany
e-mail: kresse@hs-nb.de

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_3

ogy. It starts with Sect. 3.6 about spatial databases with vector and raster models, referencing the relevant standards. Section 3.7 covers spatial queries and filtering. Section 3.8 explains indexing, which supports acceleration of queries. Section 3.9 provides an overview of network databases and some prominent network search algorithms. Section 3.10 is dedicated to raster databases, and Sect. 3.11 introduces time in the context of spatiotemporal databases. Section 3.12 summarizes the most widespread database software solutions.

Keywords

georeferencing database · Earth browser · geoconfiguration · GI storage

3.1 Historical Background

The term *database* was coined with the advent of the computer. However, databases are not a twentieth century concept. Though known under different names, databases have been applied for centuries. Early taxation systems are one of the oldest examples. Figure 3.1 shows a late seventeenth century tax-cadastral map that is linked to a table of tax-relevant parameters necessary to determine the correct farmer's tax (Table 3.1).

Today, databases are present in every corner of our life. Modern database technology was developed for administration of bank accounts, followed by warehouse management. Databases are scalable from small desktop solutions to huge server clusters, e. g., a private photograph collection versus administration of an entire cellphone system. Some databases may typically have only a small number of modifications per time period, while others are very dynamic. A telephone directory or a railroad schedule has a static nature, while an Internet shop or a monitoring system for vehicle movements is very dynamic.

Fig. 3.1 Late seventeenth century *Matrikel* map of Kröpelin/Germany, showing property boundaries, land use, and pointers to the tax data tables (adapted from [1])

3.1.1 Features of a Database

A *database* is an organized computer-based collection of data that allows the management of those data, including insertion, modification, retrieval, and deletion. The term database comprises the software and hardware involved, including the physical data storage. The operation of a database is controlled by a database management system (DBMS) that provides the interfaces for the communication from the outside and conceals the physical data storage from the applications.

A database has four typical properties [2].

Reliability

It shall deliver an uninterrupted service and be able to cope with unforeseen situations like the interruption of a data link. For example, during a purchase via the Internet, we expect that the database will be reset to its original state if the Internet-connection breaks down after we have paid with our credit card but before the purchase has been confirmed.

Correctness and Consistency

The internal logics of the data of our database shall be correct. For example, if we have created a database of our photographs, we expect that the photos shown in the table of contents exist in the database, and the other way around, the photos that we have loaded appear in the table of contents, too.

Technology Proof

The DBMS shall be independent of the details of the hardware and software. It is commonplace that computer hardware and software keep on developing fast. However, if we visit an Internet shop, we do not want to know, which software and hardware it is built of, and if this has been updated recently.

Table 3.1 Example of a tax data table

Attribute
Pointer: L4
Municipality: Kröpelin
Location of the village
Proprietor
Municipality
Name of owner
Societal position of owner (full farmer, half farmer, *cottager*) or profession (tailor, blacksmith)
Size and quality of fields
Description of meadows, pastures, woods, waters, roads
Soil quality, quantity of sowing, quantity of yield, livestock
Tax obligation
Labor service obligation

Security

The data of a database shall be protected against loss and unauthorized read and write access. The typical example regards a bank account. We do not want to let anybody else read how much money we have or our debts (read access). At the same time, the bank does not allow us to change the amount of money on our account by simply typing in new figures (write access). Typically, data access is partitioned into two or more levels of rights, such as user and administrator rights.

3.1.2 Database Architecture

A database is structured in three levels: the external level, the conceptual level, and the internal level.

A user usually has only limited access to the database. We again may think of our bank account. Such an access is called a view of the database. A view is defined by a subset of the database content and a level of authorization. The totality of all views forms the external level.

Within the database, the data are organized and stored in a way that was designed before the database was set up. The database design is explained further below. For example, consider again the database of our photographs. It may have a simple hierarchy with folders named according to events, such as holiday_2018 on the top level and the photographs on the bottom level. Such a structure is called a user-oriented conceptual schema. It resides on the conceptual level of the database.

The internal schema comprises all aspects of the physical data storage. The related level is called the internal level.

The three-level architecture was developed in the 1970s and standardized by the American National Standards Institute (ANSI)/Standards Planning and Requirements Committee (SPARC) (Fig. 3.2; [3]).

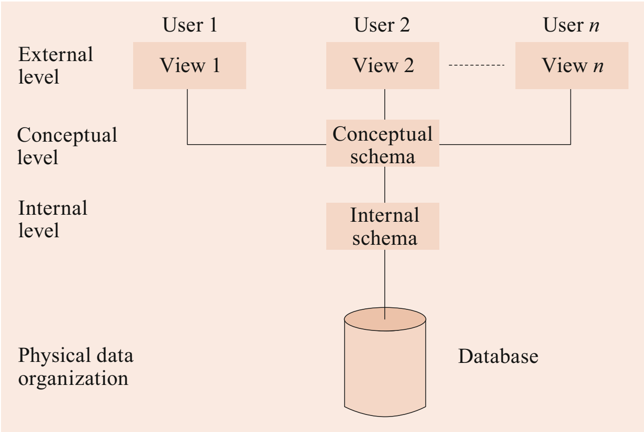


Fig. 3.2 ANSI/SPARC three-level architecture (adapted from [2, 3])

Database Management System (DBMS)

A database management system (DBMS) is the totality of software components that define the data model, realize all database properties, and provide the interface between the application programming interface (API) and the physical storage of the data.

Figure 3.3 explains the structure of and the workflow within a DBMS. A query is submitted to the DBMS through the user interface using a data interaction language, usually the Structured Query Language (SQL). The user interface is typically an API or the console for the administrator's access. The query is forwarded to the query compiler, which either sends it directly to the runtime database processor for forming the final retrieval command or to the query optimizer. The query optimizer is invoked to improve the database performance. For instance, if a query requires a Cartesian product of three tables that are significantly different in size, say 100, 10 000, and 1 000 000 rows, then it is faster to first compute the product of the two smaller tables followed by the large one, compared to computing the product of the two larger tables first.

The database may hold, for example, land parcels whose areas are constrained to values larger than zero square meters. Every time the database is modified, the constraint enforcer checks the permissibility under the defined constraints and refuses a change request when necessary. The runtime database processor handles the access to the physi-

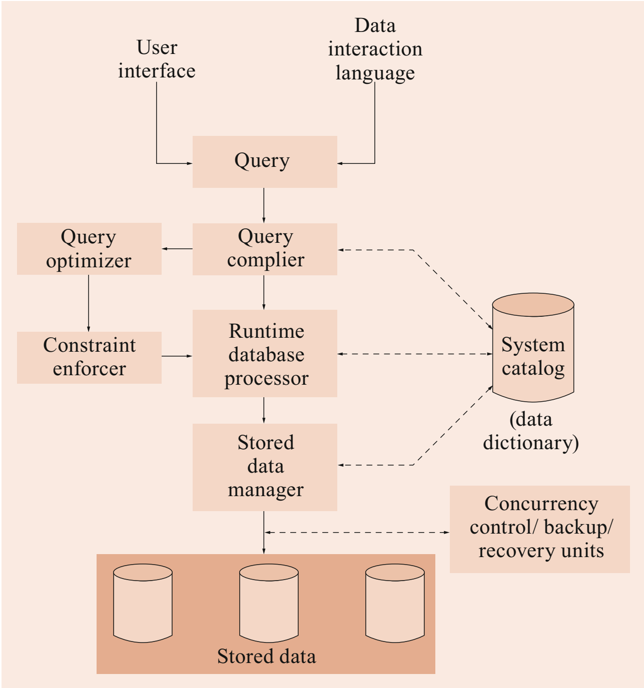


Fig. 3.3 DBMS components used to process user queries (adapted from [2])

3

Table 3.2 Lost update for nonatomic interleaved transactions, T_1 and T_2, with variables X and Y and bank balance B (adapted from [2])

T_1	T_2	B (€)	X (€)	Y (€)
		1000		
$X \leftarrow B$		1000	1000	
$X \leftarrow X + 300$ €		1000	1300	
	$Y \leftarrow B$	1000	1300	1000
	$Y \leftarrow Y - 400$ €	1000	1300	600
$B \leftarrow X$		1300	1300	600
	$B \leftarrow Y$	600		600

cal data storage, technically speaking to the files in which the data actually reside. At this level, critical situations regarding the overlapping access to the data and data preservation in the case of unforeseen disruptions are cared for. These situations are named concurrency control, backups, and recovery, which are explained below. Sometimes, the part of the DBMS that controls the critical situations is called the transaction engine. The query compiler, the runtime data processor, and the stored data manager read their control data from the system catalog, which is a data dictionary that holds the three schemas explained in Sect. 3.1.2.

3.1.3 Operational Requirements

The basic demands on a database are quite self-evident. A database query shall be answered after a reasonable time span, and the works of simultaneous operations shall not interfere. In other words: the availability and the performance shall be good. The accesses of many users shall be independent. The latter can be established by locking that part of a database to which one user has granted access.

Transactions

Above, we saw the example of a disruption of the network connection while paying for an Internet purchase. Such an irregularity must not lead to a loss of our money and, seen from the logics of the database, it must not lead to an inconsistent dataset. The payment is an example of a database transaction. Transactions are triggered with the commands *Insert*, *Modify*, *Delete*, and *Retrieve*. A transaction follows four rules.

- The transaction is *atomic*, as it cannot be split into smaller transactions. That is, the user can rely on a transaction performing all its actions in one step or performing none of the actions at all.
- The transaction is *consistent*, as it begins and ends with a consistent database.
- The transaction is *isolated*, as the involved data cannot be accessed by other users during the transaction. While interacting with the database a user should feel that he or she is the only user at that time.
- The transaction is *durable*, as it ends with a permanent modification of the database.

Because of the first letters of the four keywords, the rules are often called ACID rules.

After a successful transaction, the database is set to the new state. The related command is named *Commit*. If a problem occurs, the database is set back to its original state. This is called a recovery of the database. The related command is named *Rollback*.

The simultaneous access of many users to the same database requires an exclusion of more than one write access to the same data. This is called concurrency control. A good example of the consequences of disregarding concurrent accesses to the same data, which lead to an inconsistent database, is shown in [2].

For example, suppose that my bank balance is 1000 €. Two transactions are in progress: T_1 to credit my account with 300 €, and T_2 to debit 400 € from my account. Table 3.2 shows a particular sequence of the constituent operations of each transaction, termed interleaving. Transaction T_1 begins by reading my balance B from the database into a program variable X and increasing X to 1300 €. Transaction T_2 then starts by reading the same balance from the database into Y and decreasing Y to 600 €. T_1 then concludes by writing X to the database as the new balance of 1300 €, and T_2 writes Y to the database as the new balance of 600 €. It is as if transaction T_1 never occurred, a problem known as lost update. Interleaving can improve database performance, because shorter operations may be executed, while more lengthy operations are still in progress. However, interleaving must be controlled to avoid problems such as lost update [2].

3.1.4 Data Models

In Sect. 3.1.2, we mentioned the example of a database of photographs with folders named according to events and the photographs of those events underneath those folders. This structure can be imagined as a simple tree with the folders as the branches and the photographs as the leaves. The corresponding data model is a hierarchical model.

A data model is an explicit definition of the structure of data. In modern database technology, only two types of data models (*database models*) are common: the relational model and the object-oriented model. The relational model subdivides into the relational model by *Codd* [4] and the entity-relationship model by *Chen* [5, 6].

The object-relational database management system (ORDBMS) combines elements of the relational and the object-oriented models. The hierarchical model is applied in the simple example above. It offers efficient data storage, but its application is slightly decreasing because of the limitations. For instance, it cannot model the links between many house-owners and many insurance companies, which is the many-to-many relationship shown in Sect. 3.2.4.

The relational model is currently the most common model by far and is discussed in Sect. 3.2.

The applications of the entity-relationship model are rarer. However, after almost four decades, it is still acknowledged because of its semantic expressiveness. It is common ground that its concepts have influenced the development of the object-oriented model and probably will continue playing their role in future developments. The entity-relationship model is discussed in Sect. 3.5.

The object-oriented model overcomes many limitations of the relational model. For instance, the storage of a polygon in a relational model according to Codd is laborious, because it cannot be extended to inhabit new data types. Such limitations have been overcome by the object-oriented model, which is discussed in Sect. 3.3.

3.2 Relational Model

This section regards the relational database in the way it was defined by *Codd* in 1970 [4]. The data model is based on set theory and predicate logic. Set theory is explained in Sect. 3.2.2. Predicate logic is a branch of mathematical logic that considers the subject–predicate relation [7].

A relational database consists of a group of unordered tables. Each table has a number of rows and columns. Each row represents an item that is described by the attributes, which are placed in the columns. *Codd* called such a table a relation because it inhabits a related set of information [2]. In other words, each row provides information that is interrelated by residing in the same table. This is in contrast to the popular but wrong thinking that the name relational database refers to the relation between the tables.

Often, the term tuple is used instead of row. Originating in set theory a tuple is a sequence of n elements where n is a positive integer.

The relational database model gained wide acceptance because of its simplicity. The major benefits are quoted in [8].

- Data entry, updates, and deletions will be efficient.
- Data retrieval, summarization, and reporting can be efficiently computed by introducing techniques like indexes (Sect. 3.5).
- Since the database follows a well-formulated model, it behaves predictably.
- Since much of the information is stored in the database rather than in the application, the database is somewhat self-documenting.
- Changes to the database schema are easy to make.

3.2.1 Design

To state the obvious: a database is always an abstracted representation of the real world and, thus, can never contain any element that one could think of. The definition of the conceptual schema, the real world's subset and its structure, and the mapping to the tables of a database is called the design of the database. The design is based on the three-level-architecture explained in Sect. 3.1.2.

Unfortunately, the design of a database cannot be automated. It rather requires decisions on the shape of the future database. For example, a database is to manage the plants of a horticulture company. We assume that apple trees, oak trees, black-currant bushes, and lilac bushes exist. If we want to create two types of tables, we may put the trees in one table (apple and oak) and the bushes in the other (currant and lilac). Alternatively, we could put the fruit plants in one table (apples and currants) and nonfruit plants in the other (oaks and lilacs), which is probably the more adequate solution for a horticulture company.

3.2.2 Tables

In a relational model, a table is the central element. A table represents the real world as far as it is to be modeled for a given task. A relational database consists of one or many tables that are related to each other. Obviously, one table, or even a few, would be an exceptional case. As shown in the following examples, a table may represent the gardens/parcels of a city, the owners of those gardens/parcels, or further information about the owners, such as their telephone numbers. A table consists of rows and columns. Each row represents an instance of the real-world item that the table represents. Each column represents an attribute of the item [2, 4, 9].

There are a number of rules regarding the creation of the rows of a table in the relational model. These rules guarantee unambiguous addressing of every item of information in the relational database by programming. The uniqueness of a row within a table can be guaranteed by designating a primary key. This is a column or a combination of columns that contains unique values throughout the table. It is used to address every individual row and its information by giving the table name and the primary key. By definition, each table can

Table 3.3 Table *owner*: The best choice for primary key for *owner* would be OID

owner						
OID	LastName	FirstName	Address	ZipCode	City	Telephone
1	Rohde	Thomas	Kleeblattstraße 19	12524	Berlin	030/1234567
2	Lenz	Rolf	Blankenburger Weg 10	22459	Hamburg	040/1234567

only have one primary key, even though several columns or combinations of columns (called candidate keys) may have unique values. The primary key must be selected from one of the candidate keys.

It is up the designer of the relational database to decide which of the candidate keys is to become the primary key. The decision should be based upon the principle of [10]:

- Minimality (choose the least number of columns necessary)
- Stability (choose a key that seldom changes)
- Simplicity/familiarity (choose a key that is both simple and familiar to users).

The main properties regarding tables—relationships and normalization—are elaborated using an example featuring owners, gardens, and plants. Let us assume that a city has a table of garden owners called *owner*, which looks like the table shown in Table 3.3.

The candidate keys for *owner* include OID and Telephone. Telephone is not favorable because the number might change frequently over time.

How should we decide regarding OID and LastName + FirstName? Both choices are reasonable, although names change sometimes, for instance with marriage. However, the answer is easy, because numeric columns can be searched and sorted more efficiently than character columns.

The previous paragraphs outlined the theory. In many practical applications, the primary key is an additional static integer number that does not change over time, because it is not selected from the information given in a table. A typical primary key is an order identifier (ID) or simply a row counter.

Foreign Keys

A relational database is built of tables. Primary keys are used to create the relationships between the tables. A foreign key in one table points to a primary key in another table or, expressed the other way around, a foreign key is a column in a table used to reference a primary key in another table.

Continuing the example given in the last section let us assume that we choose OID as the primary key for *owner*. Now we define a second table, *garden*, as shown in Table 3.4.

OID is considered a foreign key in *garden*, since it can be used to refer to a given person, i. e., a row in the *owner* table [8].

Table 3.4 Table *garden*: OID is a foreign key in *garden* which can be used to reference an owner stored in the *owner* table

garden				
GID	OID	Area (m^2)	Price (€)	Soiltype
11	1	701	701	Mould clay
12	2	877	877	Loam
13	1	1305	1305	Loam

Table 3.5 The tables *garden_basic* and *garden_confidential* are related in a one-to-one relationship. The primary key of both tables is GID

garden_basic
GID
Area
Soiltype
One-to-one (1 : 1)
garden_confidential
GID
OID
Price
One-to-one (1 : 1)

3.2.3 Relationships

Relationships in the real world might be quite complex. Think of a resident in a city. He or she has numerous private and official relationships within the neighborhood, with the city council, and beyond. However, the model of a relational database only allows relationships between pairs of tables. These tables can be related in one of three different ways [8]

1. One to one (1 : 1)
2. One to many (1 : n)
3. Many to many (m : n).

One-to-One Relationships

In this simple case, two tables are related, such that each row of the first table has at most one partner row in the second table. This type of relation rarely exists in the real world. However, it is often applied in order to hide information, such as personal data, in a second table.

In the example shown in Table 3.5, the prices of the gardens are separated from the technical data, e. g., soil type.

Another example may be a large table of which only a part is needed in another application. Then it may be advisable to split the large table into two which are related with a one-

Table 3.6 There can be many trees, bushes, and vegetables in a garden, so *garden_basic* and *garden_plants* are related in a one-to-many relationship

garden_basic
GID
Area
Soiltype
One-to-many (1 : *n*)
garden_plants
GID
Trees
Bushes
Vegetables
One-to-many (1 : *n*)

to-one relation, in order to use only the smaller part-table in the other application. For the sake of clarity, tables that are related in a one-to-one relationship should always have the same primary key [8].

One-to-Many Relationships

A one-to-many relationship of two tables creates a tree-like structure. This means that, for each row in the first table, there can be zero, one, or many rows in the second table, but for every row in the second table there is exactly one row in the first table. For example, each garden can have many trees. Therefore, garden_basic is related to garden_plants in a one-to-many relationship (Table 3.6). It is obvious that the one-to-many relationships are the standard case in relational databases.

Many-to-Many Relationships

A many-to-many ($m : n$) relationship of two tables means that a row of the first table may have many related rows in the second table, and that at the same time a row in the second table may have many related rows in the first table. An example is the relation between house owners and insurance companies/their insurance programs. A house owner usually has contracts with several insurance companies, while an insurance company has many house owners as customers. Thus, the *houseOwner* table in a real-estate database would be related to the *insurer* table in a many-to-many relationship.

Many-to-many relationships cannot be modeled in relational databases and, therefore, have to be split into two steps: a one-to-many ($1 : n$) relation and a many-to-one ($n : 1$) relation. This requires an intermediate table that links to both original tables, in the example above, the house owners and the insurance companies. In the example shown in Table 3.7, this linking table is called *houseOwnerLinksInsurer* and would contain one row for each insurance program of each house owner.

Table 3.7 A linking table, *houseOwnerLinksInsurer*, is used to model the many-to-many relationship between *houseOwner* and *insurer* (after [8])

houseOwner
IDNumber
LastName
FirstName
Address
ZipCode
City
One-to-many (1 : *n*)
houseOwnerLinksInsurer
IDNumber
InsurerID
insurer
InsurerID
CompanyName
Address
ZipCode
City
One-to-many (1 : *n*)

3.2.4 Normalization

The design of a database schema allows for a number of choices. Those choices include the definition of tables which may result in a few large tables or many smaller tables. A large table with many columns may have all information in one row. A small table with a few columns may allow for a better overview and more flexible creation of relationships between the tables. The primary and foreign keys have been discussed above.

Normalization may be considered as the process of simplifying the database structure to minimize the dependency between tables and to allow for the greatest diversity of database queries. The normal forms are a linear progression of rules that you apply to your database, with each higher normal form achieving a better, more efficient design [8].

The basic normal forms are

- First normal form (1NF)
- Second normal form (2NF)
- Third normal form (3NF).
- Higher levels of normal forms, such as the Boyce–Codd normal form and the fourth normal form.

First Normal Form

The discussion of the normal forms continues and extends the garden example that has been mentioned already in Table 3.3. Table 3.8 presents all involved data of our garden example that will be restructured step by step to meet the first, second, and third normal form (1NF, 2NF, 3NF).

Table 3.8 *garden_1* violates first normal form because the data stored in the Species column are not atomic

garden_1				
GID	Parcel	Owner	Date of purchase	Species
11	110	Rohde	1980	5 Apple trees, 3 Plum trees, 6 Peach trees
12	130	Lenz	1995	1 Apple tree
13	150	Rohde	1985	2 Chestnut trees, 2 Oak trees

The first normal form (1NF) demands that all column values are atomic. This literally means that they are indivisible. Arrays of values are not allowed because it would be disadvantageous for easy update or retrieval of data.

The table *garden_1* has some attributes that contain more than one element, e.g., 5 Apple trees, 3 Plum trees, and 6 Peach trees. Thus, the table *garden_1* (Table 3.8) violates the 1NF.

It is quite obvious that it would take some programming effort to query the data such as: show me all gardens with apple trees. Query programs using tables such as *garden_1* tend to contain more programming errors than queries on simpler structured tables.

The same table could be improved by replacing the single Species column with six columns: Quant1, Specie1, Quant2, Specie2, Quant3, and Specie3 (Table 3.9), but it still violates the 1NF.

The design shown in Table 3.9 is still problematic.

1. A search for specific species would need to read through all columns, e.g., to retrieve the apple trees.
2. The number of different species in a garden is limited to three because the table offers only three pairs of columns, i.e., Quant1/Specie1 etc.

Obviously the table could be expanded to more than three species. However, where is the limit? Any limit might be too narrow for some exceptional cases. Otherwise, if the table is defined large, in most cases it would cause a waste of empty space.

A table in first normal form (1NF) does not have those problems. The table *garden_31* in Table 3.10 is in 1NF, as each column only contains one value and there are no groups of repeating columns.

However, each tuple (row) requires an unequivocal and preferably stable primary key. In order to fulfill this requirement a column SID (species ID) is added which results in table *garden_32*. The SID relates to the species. For instance *21* stands for *Apple tree*. GID and SID form the composite primary key.

The rules for 1NF can be summarized as follows [11]

- Eliminate repeating groups in individual tables.
- Create a separate table for each set of related data.
- Identify each set of related data with a primary key.

Second Normal Form

A table is considered to be in second normal form (2NF) if it is in 1NF and every nonkey column is fully dependent on the entire primary key. In other words, tables should only contain data that are related to one entity (thing) in the real world.

To find out whether table *garden_32* meets second normal form (2NF), one has to investigate whether all columns fully depend on the primary key, which is a composite of the columns GID and SID in the given case.

However, the columns *Parcel*, *Owner* and *Date of purchase* do not depend on all parts of the primary key. This leads a redundant storage of values which can be observed in table *garden_32*. Redundancy may cause anomalies with insertion, update, and deletion of records. For this reason *garden_32* is not in 2NF.

The second normal form can be achieved by splitting *garden_32* into two tables. The process of splitting a non-normalized table into its normalized parts is called decomposition. Since *garden_32* has a composite primary key, the decomposition process is easy: one simply puts everything that applies to each garden in one table, and everything that applies to each species in the second table. The two decomposed tables in second normal form, *garden_4_garden* and *garden_4_species*, are shown in Table 3.11.

Two thoughts should be considered.

During the normalization process, no information is deleted. This form of decomposition is termed nonloss decomposition, because no information is sacrificed to the normalization process.

The decomposed table can be recombined to one table. This is guaranteed by the common foreign key, in the example GID.

Table 3.9 *garden_2*: A better, but still erroneous, version of the gardens table. The repeating groups of information violate first normal form

garden_2									
GID	Parcel	Owner	Date of purchase	Quant1	Specie1	Quant2	Specie2	Quant3	Specie3
11	110	Rohde	1980	5	Apple tree	3	Plum tree	6	Peach tree
12	130	Lenz	1995	1	Apple tree				
13	150	Rohde	1985	2	Chestnut tree	2	Oak tree		

Table 3.10 *garden_31* table: in first normal form but no proper primary key. *garden_32* table: in first normal form with composite primary key: GID and SID

garden_31					
GID	Parcel	Owner	Date of purchase	Species	Quantity
11	110	Rohde	1980	Apple tree	5
11	110	Rohde	1980	Plum tree	3
11	110	Rohde	1980	Peach tree	6
12	130	Lenz	1995	Apple tree	1
13	150	Rohde	1985	Chestnut tree	2
13	150	Rohde	1985	Oak tree	2

garden_32						
GID	SID	Parcel	Owner	Date of purchase	Species	Quantity
11	21	110	Rohde	1980	Apple tree	5
11	22	110	Rohde	1980	Plum tree	3
11	23	110	Rohde	1980	Peach tree	6
12	21	130	Lenz	1995	Apple tree	1
13	24	150	Rohde	1985	Chestnut tree	2
13	25	150	Rohde	1985	Oak tree	2

Table 3.11 Set of two tables in second normal form: *garden_4_garden* with primary key GID and *garden_4_species* with primary key being a composite of GID and SID

garden_4_garden			
GID	Parcel	Owner	Date of purchase
11	110	Rohde	1980
12	130	Lenz	1995
13	150	Rohde	1985

garden_4_species			
GID	SID	Species	Quantity
11	21	Apple tree	5
11	22	Plum tree	3
11	23	Peach tree	6
12	21	Apple tree	1
13	24	Chestnut tree	2
13	25	Oak tree	2

Table 3.12 The *garden_4_SID* and *species* tables are in third normal form (3NF). The SID column in *garden_4_SID* is a foreign key referencing *species*

garden_4_garden			
GID	Parcel	OID	Date of purchase
11	110	Rohde	1980
12	130	Lenz	1995
13	150	Rohde	1985

garden_4_SID		
GID	SID	Quantity
11	21	5
11	22	3
11	23	6
12	21	1
13	24	2
13	25	2

LUT_species	
SID	Species
21	Apple tree
22	Plum tree
23	Peach tree
24	Chestnut tree
25	Oak tree

The rules for 2NF can be summarized as follows [11]

- Create separate tables for sets of values that apply to multiple records.
- Relate these tables with a foreign key.

Third Normal Form

A table is said to be in third normal form (3NF) if it is in 2NF and if all nonkey columns are mutually independent.

The dependency that violates the 3NF may be of various kinds, for instance:

- A new column is derived from two others. An example could be the price for a garden which is computed by the product of the area (in one column) and the price per m^2 (in a second column). If this price is stored in a third column, then that column depends on the other two and must be updated if the value in one of the others is changed.
- Two columns have the same content but are simply coded in a different way. In Table 3.11, table *garden_4_species*, the value 21 in the column SID is exchangeable with the value Apple tree in the column *Species*. So this is redundant information in the table. A development towards the third normal form (3NF) eliminates this redundancy.

In our example, the table *garden_4_species* can be further decomposed to achieve 3NF by breaking out the SID–Species dependency into the lookup table *LUT_species*, as shown in Table 3.12. As a result we have the new table *garden_4_SID* and the lookup table *LUT_species*. The column SID is duplicated to appear in both tables, i.e., as a foreign key in *garden_4_SID* and as a primary key in LUT_*species*. This allows an easy join of the two tables using a query.

The rule for 3NF can be summarized as follows: Eliminate fields that do not depend on the primary key [11].

Higher Normal Forms

Besides the first through third normal form, higher normal forms have been defined. The Boyce–Codd normal form (BDNF) adds one more requirement to the third normal form: Every determinant must be a candidate key.

The fourth normal form (4NF) also adds one more requirement to the third normal form: A relation is in 4NF if it has no multivalued dependencies. A multivalued dependency happens if two attributes (or columns) in a table are independent of one another, but both depend on a third attribute. The normalization standard fourth normal form (4NF) prevents a multivalued dependency [12].

3.2.5 Integrity Rules

The integrity rules are a part of any relational database and formally express the consistency that is one of the database properties as shown in the beginning of this chapter. Two types of integrity rules can be distinguished: general and database specific.

General Integrity

The two general integrity rules are

- Entity integrity
- Referential integrity.

The entity integrity rule says that primary keys cannot contain Null (missing) data. The reason is obvious, as one cannot uniquely reference a row in a table, if the primary key could be Null. This rule also applies to composite keys, where none of the individual columns may be Null.

The referential integrity rule says that the database must not contain any unmatched foreign key values. This implies that [8]

- A row may not be added to a table with a foreign key unless the referenced value exists in the referenced table.
- If the value in a table that is referenced by a foreign key is changed (or the entire row is deleted), the rows in the table with the foreign key must not be *orphaned*.

Generally speaking, three options exist to overcome the problem:

- *Disallow*. The change is completely disallowed.
- *Cascade*. For updates, the change is cascaded to all dependent tables. For deletions, the rows in all dependent tables are deleted.
- *Nullify*. For deletions, the dependent foreign key values are set to Null [8].

Database-Specific Integrity

All other integrity constraints are database-specific rules. They depend on the specific application that is to be served by the database [8].

If we follow the garden example above, the following constraints may apply.

- The zip code must have exactly five digits.
- The house number must be greater than zero.
- The area of a garden is never negative.

3.2.6 Set Theory

The set theory is a branch of mathematics that studies sets. A set is any collection of distinct entities of any sort. Examples are $A = \{1, 2, 3\}$, $B = \{a, b\}$, and $C = \{\text{Bach, Haydn, Mozart}\}$.

Part of Codd's genius was to recognize that many of the standard set operators map nicely to real data manipulation problems.

Set-Based Operations

There are six fundamental relational operators, namely

- Union,
- Difference
- Product
- Project
- Restrict
- Rename.

The first three of these are traditional set-based operators.

Further relational operators are

- Intersection
- Join.

They are termed derived relational operators and can be expressed using different combinations of the six fundamental operators.

The structure of these operations and the way in which they can be combined is called relational algebra. In Sect. 3.2.7, it will be shown how the SQL is built upon the relational algebra.

The following listing addresses the operations in detail.

- The union operation is a binary operation that takes two sets and returns the set of elements that are members of at least one of the original sets. The union of sets A and B is denoted $A \cup B$ and defined by $A \cup B = \{x | x \in A \vee x \in B\}$. Example: if $A = \{1, 2, 3\}$ and $B = \{1, 2, 4, 5\}$, then $A \cup B = \{1, 2, 3, 4, 5\}$.

- The difference operation is a binary operation that takes two sets and returns the set of elements that are members of the first set but not the second set. The difference of sets A and B is denoted $A - B$ and defined by $A - B = \{x | x \in A \wedge x \notin B\}$. Example: if $A = \{1, 2, 3\}$ and $B = \{1, 2, 4, 5\}$, then $A - B = \{3\}$.
- A Cartesian product is a set of ordered pairs, produced by combining each element of one set with each element of another set, denoted by $A \times B$. Example: let $A = \{1, 2, 3\}$ and $B = \{a, b\}$, then $A \times B = \{\langle 1, a\rangle, \langle 1, b\rangle, \langle 2, a\rangle, \langle 2, b\rangle, \langle 3, a\rangle, \langle 3, b\rangle\}$. Note that Cartesian products may be generated by multiplying any number of sets together. The actual number of sets involved in a particular case is said to be the *degree* or *arity* of that Cartesian product. In this example the degree is two. A relation is a subset of a Cartesian product. An example is $Q = \{\langle 1, a\rangle, \langle 1, b\rangle\}$. Note that relations may be also of any degree or arity [9].
- The project operation is unary, applying to a single relation. It returns a new relation that has a subset of attributes of the original. The relation is then modified so that any duplicate tuples formed are coalesced.
- The restrict operation is also unary. The restrict operator works on the tuples of the table rather than the columns and returns a new relation that has a subset of tuples of the original. A condition specifies those tuples required. The restrict operator is often referred to as the select operator.
- The rename operation is unary and defined as $\rho_{a/b}(R) = \{t[a/b] : t \in R\}$ Example: $\rho_{\text{house/building}}(\text{City})$ renames the column house to building in the relation (table) City.
- The intersection operation is a binary operation that takes two sets and returns the set of elements that are members of both the original sets. The intersection of sets A and B is denoted $A \cap B = \{1, 2\}$, being the set defined as $A \cap B = \{x | x \in A \wedge x \in B\}$. Example: If $A = \{1, 2, 3\}$ and $B = \{1, 2, 4, 5\}$, then $A \cap B = \{1, 2\}$.
- Join is a binary operator that takes two relations as input and returns a single relation. The join operation allows connections to be made between relations.

Figure 3.4 shows by shading on set diagrams some of the set operations described above. The application of the set theory to the Structured Query Language is shown in Sect. 3.2.7.

3.2.7 Structured Query Language

The standard method of communicating with relational databases is SQL. In contrast to most other languages, it is neither a programming language nor a markup language.

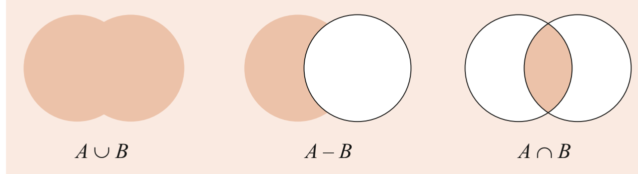


Fig. 3.4 Set union, difference, and intersection (adapted from [2, 4, 9])

SQL may rather be called a *set language*, as it implements set theory. The term *relational algebra* is another way of expressing computations based on the set theory.

SQL was developed by IBM Research in the mid 1970s, completed as a US American National Standards Institute (ANSI) standard in 1986, and adopted as an ISO standard in 1987. ISO/IEC 9075-*:1999, also known as SQL:1999 or SQL3, added some object-oriented features, as discussed below. The latest version is ISO/IEC 9075-*:2016.

The relational model defines two *root* languages for accessing a relational database: relational algebra and relational calculus. Relational algebra is a low-level, operator-oriented language. Creating a query in relational algebra involves combining relational operators using algebraic notation. Relational calculus is a high-level, declarative language.

The basic structure in SQL is the statement. Semicolons separate multiple SQL statements.

There are three basic categories of SQL statements [13, 14].

1. SQL-Data Statements – query and modify tables and columns
   - SELECT – query tables and views in the database
   - INSERT – add rows to tables
   - UPDATE – modify columns in table rows
   - DELETE – remove rows from tables
2. SQL-Transaction Statements – control transactions
   - COMMIT – commit the current transaction
   - ROLLBACK – roll back the current transaction
3. SQL-Schema Statements – maintain schema (catalog)
   - CREATE TABLE
   - CREATE VIEW
   - DROP TABLE
   - DROP VIEW
   - GRANT – grant privileges on tables and views to other users
   - REVOKE – revoke privileges on tables and views from other users.

Set theory being the foundation of the SQL leads to the fact that almost all operations defined by the set theory can be performed by SQL commands. Thus, the seven set-based operations (union, difference, product, project, restrict, intersection, and join) can be expressed by SQL statements. Table 3.13 provides an overview. For details, the literature may be consulted [13–15].

Table 3.13 Comparison of set-based operations and SQL statements (adapted from [15])

Description	Example	
	Structured Query Language	Explanation
Union		
A *relational union* is the concatenation of two tables	SELECT CompanyName AS Name, Address, City, PostalCode FROM Customers UNION SELECT FirstName ||' '|| LastName AS Name, Address, City, PostalCode FROM Employees ORDER BY name;	This statement selects the columns CompanyName, Address, City, and PostalCode from the table Customers and the columns FirstName, LastName, Address, City, and PostalCode from the table Employees and orders the output according to the output-column Name. The output-column Name contains the values of the column CompanyName, as well as the concatenated values of the columns FirstName and LastName
Difference		
The *relational difference* of two tables is the records that belong to one table but not the other.	SELECT Table1.* FROM Table1 LEFT JOIN Table2 ON (Table1.CustomerID = Table2.CustomerID) AND (Table1.CompanyName = Table2.CompanyName) WHERE (Table2.CustomerID IS NULL); Alternative version supported by the SQL-standard: SELECT * FROM Table1 EXCEPT SELECT * FROM Table2;	This statement selects the customers that are only in Table1 but not in Table2, i. e., the *orphans*
Cartesian product		
Like its counterpart in set theory, the *Cartesian product* of two tables combines every record in one set with every record in the other.	SELECT CustomerName, ServicePersonName FROM Customer, Service; Alternative version only supported by SQL-92: SELECT CustomerName, ServicePersonName FROM Customer CROSS JOIN Service;	This statement combines every entry in the column CustomerName of the table customer with every entry of the column ServicePersonName of the table Service
Projection		
Projection takes a vertical slice. It returns only a subset of the fields in the original table. SQL performs this simple operation using the ⟨ fieldList ⟩ section of the SELECT statement by only including the fields that you list.	SELECT LastName, FirstName, Extension FROM Employees ORDER BY LastName, FirstName;	This statement selects from the table Employees the columns LastName, FirstName, and Extension, orders the output by the LastName and by the FirstName, and thus creates a telephone directory
Restriction		
The restriction operator returns only those records that meet the specified selection criteria.	SELECT * FROM Employees WHERE LastName = 'Maier';	This statement selects from the table Employees only those rows which have *Maier* in the column LastName, and thus returns all *Maiers* of the company
Intersection		
The *intersection operator* returns the records that two tables have in common.	SELECT Table1.* FROM Table1 INNER JOIN Table2 ON (Table1.CustomerID = Table2.CustomerID) AND (Table1.CompanyName = Table2.CompanyName); Alternative version supported by the SQL-standard: SELECT * FROM Table1 INTERSECT SELECT * FROM Table2;	This statement selects the duplicate customers of Table1 and Table2
Inner join		
Join creates a Cartesian product of two tables. The *inner join* creates a simple Cartesian product and is implemented using the JOIN clause of the SELECT statement. The inner join could also be implemented by using FROM and WHERE.	SELECT Orders.OrderID, Orders.CustomerID, OrderDetails.ProductID FROM Orders INNER JOIN OrderDetails ON Orders.OrderID = OrderDetails.OrderID WHERE Orders.OrderID = 10248;	This statement selects from the tables Orders and OrderDetails all Orders with the OrderID = 10248. Therefore, it is the inverse action to that shown in Sect. 3.2.3

Table 3.13 (Continued)

Description	Example	
	Structured Query Language	Explanation
Outer join		
An outer join returns all the records returned by an inner join, *plus* all the records from either or both of the other table. The missing (*unmatched*) values will be null. Outer joins are categorized as being left, right, or full, depending on which additional records are to be included.	SELECT * FROM X LEFT OUTER JOIN Y ON <condition> ;	This statement returns all rows from X and only those rows from Y where the <condition> evaluates to True

The large objects (LOBs) are one of the extensions of the relational model based on the set theory. SQL:1999 defines two new data types: *binary large object* (BLOB) and character large object (CLOB).

These LOBs are references (or pointers) to data stored outside of the database. They are used for images, sound files, application executables, etc.

The database does not know anything about the internal structure of the LOB and so it cannot index, sort, or ensure the type safety of LOBs, unlike built-in or user-defined data types [16].

3.2.8 Entity–Relationship Model

The entity–relationship model (E–R model) is often utilized as the semantic model for the database design. Often being considered as a refinement of the relational model, the E–R model is particularly suitable for developing relational database schemas. The model was developed by *Chen* [5] in the 1970s. Though the relational data model developed by Codd gained a much wider acceptance because of its simplicity, the E–R model has a stable community of supporters. While Codd's model is build upon interrelated tables only, the E–R model allows explicit modeling of the relations between the entities. The entities are similar to objects in other models.

Sometimes the entities are called the nouns or things, and the relations are called the verbs.

3.2.9 Entities

An entity is an abstraction of a real world's phenomenon. Examples of entities are residential buildings and house owners. An entity has a type that characterizes the entity and supports the grouping of entities. Examples of entity types are buildings and persons. Entities may have attributes that store the properties of an entity. Examples are the address of the building and the name of the owner. Overall, every entity has a unique identification established by identifier assigned to each entity.

The graphical notation of entities was never standardized, which led to several graphical flavors. Chen proposed for an entity a rectangular box. The examples in this chapter follow *Chen*'s notation that was also utilized in [2]. The crow's foot notation utilized in Chap. 24 is one of the alternative notations.

3.2.10 Relations

Relationships are the connections between entities or entity types. A relationship has a relationship type and its own attributes. Therefore, the modeling of a relationship and an entity is similar. In Chen's notation a relationship is drawn as a lozenge (rhombus).

For example, let us consider the highway between Berlin and Hamburg as an entity, and along that road the town Ludwigslust as another entity. The town Ludwigslust has a relationship to that road with attributes such as the distance from Berlin (Fig. 3.5).

The E–R model distinguishes between one-to-one, one-to-many, and many-to-many-relationships. Examples are one town and one mayor, one town and many houses, and many towns and many train lines, respectively.

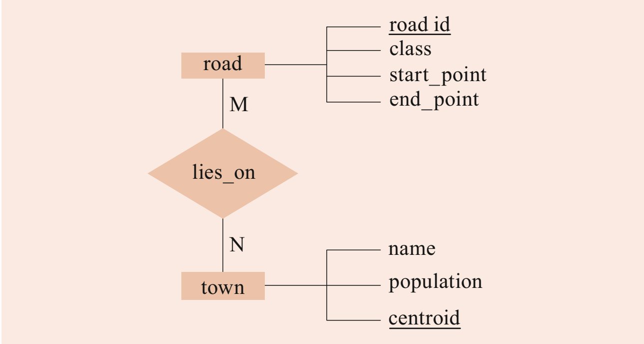


Fig. 3.5 Two entries and a many-to-many relationship (adapted from [2])

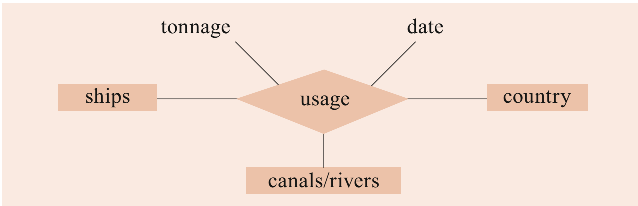


Fig. 3.6 A ternary relationship (adapted from [2])

As a special case, the E–R model defines a ternary relationship (Fig. 3.6). The name indicates a relationship of three entities, where entity 1 has a relationship to entity 2, entity 2 to entity 3, and entity 3 to entity 1. For example, consider ships, canals/rivers, and countries. The ships travel on the canals or rivers. The canals or rivers run or flow through the countries, and each ship is domiciled in one of the countries.

3.2.11 Difference to Other Models

Relational Model

Both the relational model and the E–R model utilize the Cartesian product to retrieve stored data. However, a relation after *Codd* [4] describes the structure of the *data values* within a table, while a *relationship* describes the structure of *entities* [2]. The E–R model allows modeling of rich semantics, while the relational model cannot do so. In the E–R model, the linkage between entities is explicit, while it is implicit in the relational model.

Object-Oriented Concepts

> It is commonly acknowledged that one major component of object-oriented (OO) analysis techniques is based on the E–R concepts. However, the relationship concept in OO analysis techniques is still hierarchy oriented and not yet equal to the general relationship concept advocated in the E–R model [2].

Summarizing, though the simplicity of the relational model did not fulfill all requirements, the gaps have not been filled by the E–R model that was under development at about the same time but rather by the object-oriented model that came two decades later and borrowed important techniques from the E–R model.

3.3 Object-Oriented Model

The modeling of geographic information is based on the object-oriented model. This modeling has been standardized by ISO/TC 211 and elaborated in Chap. 1.

3.3.1 Motivation for the Object-Oriented Model

While the relational database model has become extremely popular, a number of specific applications had requirements that could not be fulfilled with relational databases. Computer-aided design (CAD), computer-aided manufacturing (CAM), multimedia, imagery, and geographic information are examples of those specific applications.

All cases require new data types that are not allowed inside a column of a relational database table. New data types include sets, lists, and tables [2]. A typical case is a polygon that is a list of points. Unless a polygon cannot be stored as one data item, the software may be prone to errors, and the performance of the database low.

The object-oriented model is best suited to overcome those limitations.

3.3.2 Foundations

An object is the central element of the object-oriented model. An object can be compared to a tuple of the relational model. However, an object has a static state and a dynamic behavior. The behavior is the totality of responses to messages with its environment, i. e., with other objects. For instance, a parcel of the property cadaster may be modeled as an object with a set of operations offering read and write access to the attributes, such as size, owner, and alike. The object parcel responds this way to a database request. Often the operations are called *methods*. The object *parcel* may also store its lineage, i. e., its attributes and their change over time.

It can be stated

$$\text{object} = \text{state} + \text{behavior}.$$

Objects of similar behavior are grouped in classes. The object-oriented model offers a rich toolset and can adapt to reality much better than the relational model.

3.3.3 Features

The concept of an object as a summary of its state and behavior leads to the typical features of an object-oriented database, such as reduction of complexity, promotion of reuse, and its *metaphorical power* [2].

The size of an object can be chosen as appropriate for the model. An object may be as small as a corner stone or as large as the whole Earth. In both cases, the work with the object masks out the details of its inside, thus reducing complexity. The definition of objects often leads to the programming of software to serve the needed functionality. Typically, later projects have similar requirements compared to the earlier work. Object-oriented modeling often creates a treasure of objects, software components, that can be reused later without or with minor modifications. The term metaphorical power addresses the fact that an object – state and behavior – has a better expressiveness than a set of tables.

3.3.4 Object-Oriented Constructs

The object-oriented approach is not only one of the relevant data models for databases but also the generally accepted and standardized way for modeling geographic information. The modeling is discussed in detail in Chap. 1. For this reason, the brief characterization of the object-oriented constructs that is given in the following essentially addresses the same topic as that in Chap. 1, however, with a specialization for databases.

This section explains identity, encapsulation, inheritance, polymorphism, and association. The association has the aggregation as a subtype.

Objects are created and stay until they are deleted at some time. During the time span in between the objects are never altered. Only their attributes may change. An object possesses a unique *identity*. A parcel is created, for instance, by partitioning a larger parcel. During the parcel's lifetime, the owner or the land use may change, but the parcel stays as an object.

As discussed above, objects reduce complexity by masking out the details inside. This masking is called *encapsulation*. Everything within the object, such as its attributes, can only be accessed through well-defined interfaces. Typically, the object provides a read method and a write method for each of its attributes.

The object-oriented approach includes the generalization–specialization relationship. The generalization helps to avoid duplicate definition of properties. For instance, a geographic dataset may contain residential buildings and office buildings. Obviously, both types of buildings have attributes in common, such as footprint and number of floors, while other attributes are distinct. Both building types may be generalized to a general class Building, with attributes such as footprint and number of floors, while the specialized classes OfficeBuilding or ResidentialBuilding hold specific attributes, such as the names of the companies that rent the office space. If such an office building is to be processed, e. g., displayed on a map, the attributes are read from the class Office Building and, if necessary, from the class Building. The delivery of attributes from the class Building to the class OfficeBuilding is called *inheritance*.

Polymorphism literally means *many forms*. A simple example regards a geometric figure, such as a circle, a triangle, or a square. An object may have the behavior of displaying one of those figures on the screen. Depending on way the object is invoked, the figure will be one of the shapes listed above. The construct helps to develop more flexible software, as it splits the code into one object that is dedicated to drawing the figure and may take a default if no further information is available, and another object that is programmed to take the order, generate the specific shape, and send it out to be displayed [17].

An *association* is a grouping of objects, for instance a building and a road. This represents a very general relationship. A more specific grouping is expressed by the aggregation, as it denotes a part whole relationship. Consider the farm example given in Chap. 1. The farm consists of buildings and subareas. The buildings and subareas are aggregated to the farm. An aggregation can be organized hierarchically. For instance, the building can be modeled as an aggregation of walls, windows, doors, etc.

A hierarchy of aggregation relationships is termed a partonomy or, sometimes, a mereological hierarchy, in contrast to an inheritance hierarchy, also termed a taxonomy [2].

3.3.5 Manifesto

While a relational database system has the clear specification given by *Codd* [2, 4], no such specification existed for object-oriented database systems (OODBS) [18], even when there were already products on the market. Consideration of the features of both object-oriented systems and database management systems has led to a definition of an object-oriented database, which was presented at the First International Conference on Deductive and Object-Oriented Databases in the form of a manifesto in 1989. This manifesto distinguishes between the mandatory, optional, and open features of an object-oriented database [19].

3.3.6 Comparisons of OODBS and RDBS

The fundamentally different structure of the relational model and the object-oriented model leads to a set of difficulties that prevent a smooth transition between both models. This is usually called *impedance mismatch*. For example, in the re-

Table 3.14 The correspondence between object-oriented and relational database systems (adapted from [19])

OODBS	RDBS
Object	Tuple
Instance variable	Column, attribute
Class hierarchy	Database scheme (is-a relation)
Collection class	Relation
OID (object identifier)	Key
Message	Procedure call
Method	Procedure body

lational model, two tuples that have the same attribute value are equal. However, in the object-oriented model, the two tuples would become two objects, and two objects are never the same.

Table 3.14 shows the correspondence between object-oriented and relational database systems (RDBS).

Note that this correspondence table is only an approximate equivalence. The properties in ODBS are actually not applicable in RDBS and vice versa.

Although there are great advantages of using an object-oriented database management system (OODBMS) over a relational database management system (RDBMS), some disadvantages do exist. Table 3.15 shows the advantages and disadvantages of using ODBS over RDBS.

Originally the lack of agreed upon standards and a universal query language were some of the disadvantages. However, the Object Data Management Group (ODMG), the standardization group for the OODBMS, published the object query language (OQL), a query language that is closely related to SQL.

3.3.7 Object Relational DBMS (ORDBMS)

An ORDBMS is an attempt to extend relational database systems so that a bridge can be made between the object-oriented and relational paradigms.

The ORDBMS was created to handle new types of data, such as audio, video, and image files, that relational databases were not equipped to handle. In addition, its development was the result of the increased usage of object-oriented programming languages, and the large mismatch between these and DBMS software.

One advantage of ORDBMS is that it allows organizations to continue using their existing systems, without having to make major changes. A second advantage is that it allows users and programmers to start using object-oriented systems in parallel [20].

SQL:1999 defines a standard for object-relational extensions to SQL-92; it maintains backwards compatibility, which has placed a number of restrictions on the OO enhancements.

An object-relational DBMS provides a wider range of data types than CHAR, VARCHAR, NUMBER, etc., available in an RDBMS.

- User-defined data types: the user can define their own, more complicated data types.
- Inheritance of data types: as the number of data types increases, it is important to take advantage of the commonality between different types.
- Object identity: this enables linkage between objects and to reference one object's data from another [16].

3.4 NoSQL

Though relational databases are very widely accepted the relational model together with the structured query language is not always the ideal solution. Talking about huge datasets, the performance of relational databases sometimes tends to slow down. This disadvantage became noticeable in particular with some modern applications. They include indexing of a large number of documents, serving pages of high-traffic websites, and delivering streaming media [21].

Several additions to SQL have been developed to overcome those limitations. This set of technologies beyond SQL is termed "NoSQL" or more precisely "not only SQL". In this sense NoSQL is neither an alternative model to SQL nor an additional set of commands supplementing the SQL-standard. It is, however, a general term that summarizes a variety of technologies developed to serve a heterogeneous field of specific applications. It does not mean that SQL is replaced.

Technically spoken, NoSQL may

1. not require a fixed table schema. This means that the tables do not have a strict structure.

Table 3.15 Advantages and disadvantages of the use of OODBS (adapted from [19])

Advantage of ODBS	Disadvantage of ODBS
• Complex objects and relations • Class hierarchy • No primary keys • One data model • High performance on certain tasks • Less programming effort because of inheritance, reuse, and extensibility of code	• Schema change (creating, updating) is not trivial; it involves a system-wide recompile • Language dependence: tied to a specific programming language • Only elementary support of transaction management, query optimization, etc.

2. Usually, NoSQL avoids join operations. This means they avoid combining records from multiple tables based on common values such as primary keys.
3. NoSQL databases are typically scaled horizontally, meaning that they are scaled simply by adding more hardware.

In a more general sense, SQL/relational databases and NoSQL can be considered as subsets of structured storage. Seen from another perspective, NoSQL databases may be categorized by "data store type" such as "document store", "key/value store", "tabular" etc. [21].

NoSQL databases are applied for a large variety of purposes, including the management of geospatial data.

3.5 Indexing with B-Trees

The querying of large databases can become very time consuming. An index is a dynamic table of contents that supports fast access to datasets. Usually, an index is a tree-like structure that aids to avoid a sequential read of all possible elements of a dataset but transverses through the hierarchy of the search tree in order to find the query results a lot faster. An index can be defined for an individual attribute, for a group of attributes, or also on the basis of expressions that perform a computation on the basis of attribute values. Relational databases typically use B-trees for their search procedures.

A B-tree of order m is a search tree with the following characteristics.

1. Each node contains at most $2m$ entries.
2. Each node except for the root contains at least m entries.
3. A node with x entries as internal nodes has exactly $x + 1$ direct descendants (child nodes), or as leaf node, no descendants. The number of direct descendants is called the fan out of the node. Thus, the maximum fan out is $2m + 1$.
4. All leaves are on the same level, i. e., the distance from the root to all leaves is the same.

Figure 3.7 shows a B-tree of order 2 as an example. In order to reach optimum performance the boughs of the search tree should have similar branching, because otherwise a search in a deep-search bough may take considerably longer than in a short bough. Such trees are called balanced trees. B-trees are balanced trees, as the height h of a B-tree obeys

$$h \leq \log_{m+1} n .$$

With 1 000 000 datasets, a B-tree of order 32 reaches a maximum height of 4. In this case, maximally four accesses to database blocks are necessary to find a queried dataset.

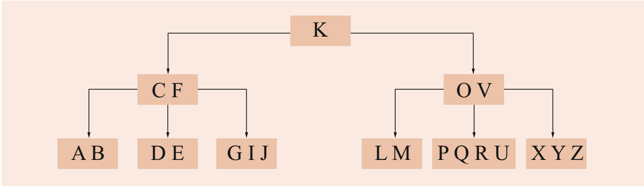


Fig. 3.7 Example of a B-tree of order 2

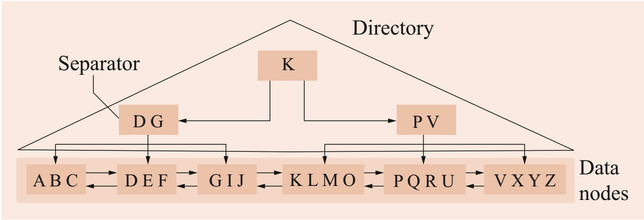


Fig. 3.8 Example of a B$^+$-tree

Because each node (with the exception of the root) is at least half-filled with data records, the storage space utilization is at least 50%.

3.5.1 B$^+$-Trees

A variant of B-trees is the B$^+$-tree (*B-plus-tree*), which stores data only in its leaves. These leaves are also called data nodes or data blocks.

The inner nodes of a B$^+$-tree form the directory. They are also called directory nodes or directory blocks. The entries in the directory nodes are called separators. For a separator sep the following search tree characteristics are valid.

$\text{sep}_l <$ sep for all separators sep_l in the *left* partial tree of sep.
S(l) $<$ sep for all datasets l in the leaves of the *left* partial tree of sep.
$\text{sep}_r >$ sep for all separators sep_r in the *right* partial tree of sep, and
S(r) $\geq$ sep for all datasets r in the leaves of the *right* partial tree of sep.

Stated differently, the directory of a B$^+$-tree is a B-tree organizing separators that in each case separate the values of two neighboring data nodes. Figure 3.8 illustrates these characteristics. This way, all separators and record keys are in the left-hand-side subtree of K, in accordance with the alphabetical sorting key smaller than K. The separators P and V in the right-hand-side subtree of K must be larger than K, as the directory is a B-tree. K is a separator between the two data nodes (G I J) and (K L M O). Similarly, V is a separator between the data nodes (P Q R U) and (V X Y Z).

Additionally, the neighboring data nodes of a B^+-tree are interlinked through appropriate references to support range queries (e. g., *Determine all places whose name starts with F to L.*). In this case, the B^+-tree serves to determine the first dataset within the queried range. Then, the data nodes are queried from left to right, until the first dataset outside of the query interval or the end of the list is encountered.

3.6 Spatial Databases

The management of spatial data is an essential task for all geographic information systems, as well as for other applications that process geospatial data. Database systems that support handling of spatial data are called *spatial database systems*. They enable storage of the data in a *spatial database* and provide mechanisms for convenient modeling and efficient querying of spatial data. Spatial database systems allow the formulation of powerful queries by providing geometric and topological operations [7, 22, 23].

The representation and querying of geometric data types exceed the capabilities of the relational database model; for example, the boundary of a polygon is typically represented by a chain of line segments. In addition, the polygon may have an arbitrary number of holes. The first normal form would require the elements of such a polygon to be distributed to several tables with corresponding relationships. Thus, a user must keep in mind the storage model while performing a query, which is user-unfriendly and violates the data independence principle. Furthermore, such spread storage would lead to rather inefficient access and costly reconstruction of a polygon. Geometric functionalities, such as a point-in-polygon test, cannot be expressed by the capabilities of the relational database model either. In addition, the indexing mechanisms of relational database systems are not sufficient for indexing spatial data because of the higher dimensionality of geometries.

There are two different approaches to achieving this objective: (1) The object-oriented or (more often) the object-relational database model is used as the basis for a spatial database. These database models offer suitable modeling capabilities and extension mechanisms for adding the required geometric data model, analysis functions, and query capabilities. (2) The geometries are stored in BLOBs using a standardized binary representation. Analysis and query functionalities are added by procedural extensions. Both approaches require the support of spatial indexing by the internal level of the DBMS.

The most important data models for spatial databases are presented in this section. Spatial query techniques and indexing structures, as well as special spatial databases (such as raster and network databases) are topics of the following sections.

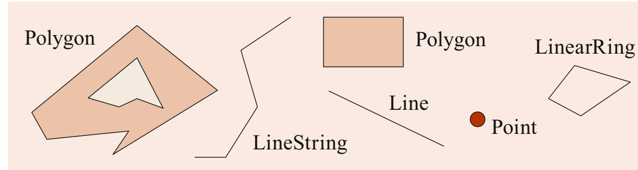


Fig. 3.9 Examples of simple features

3.6.1 Geometry Model

By offering open, standardized, and integrated storage of alphanumerical and spatial data, spatial databases facilitate data exchange and interoperability between different spatial and nonspatial applications. Two important standardized data models for representing vector data in spatial databases exist: the ISO standard 19125:2004 *Simple Feature Access* and the ISO/IEC standard 13249-3 *SQL/MM Spatial* (Chaps. 15, 30).

ISO 19125 (2-D)

ISO 19125, part 1, defines a data model for simple features. *Simple features* are features restricted to a two-dimensional geometry with linear interpolation between vertices. Figure 3.9 shows some examples of such geometries.

The data model of ISO 19125 defines a hierarchy of classes representing simple-feature geometries: The class *Point* describes zero-dimensional geometries. Curves are instantiated by *LineString* (a curve with linear interpolation between its points), *Line* (a line segment), and *LinearRing* (a closed LineString) objects. Surfaces can be represented by the *Polygon* class that allows the description of polygons with holes using linear rings. Furthermore, collection classes exist for geometries consisting of primitives of the same type (*MultiPoint*, *MultiLineString*, and *MultiPolygon*) and of different types (*GeometryCollection*). The classes provide

- Access methods to the properties and components of the geometries.
- Basic methods, such as the computation of the minimum bounding rectangle or well-known text (WKT) and well-known binary representations (WKB).
- Methods for testing topological predicates (such as intersects, contains, and touches) between geometric objects.
- Spatial analysis methods (e. g., distance, buffer, intersection, and union).

Figure 3.10 illustrates the class model of ISO 19125. It was derived from and is common with the Open Geospatial Consortium (OGC) *OpenGIS Implementation Specification for Geographic Information – Simple feature access*. Products may claim *implementation* of one or the other, or both, but only products that have been certified by the OGC can be called *OGC compliant*.

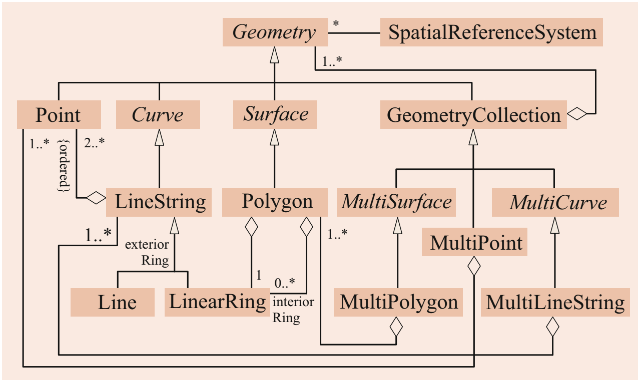


Fig. 3.10 Class model of ISO 19125-1:2004

SQL/MM Spatial (2-D)

SQL/MM Spatial is a part of the SQL/MM standard that specifies the multimedia extensions of SQL. ISO/IEC 13249-3:2016 SQL/MM Spatial extends the ISO 19125 data model; the essential part of the class model is depicted in Fig. 3.11.

The main extension in the data model is the support of circular interpolations. The class *ST_CircularString* represents curves that possess only circular interpolations, and the class *ST_CompoundCurve* allows the description of curves that have linear as well as circular interpolations. Objects of the class *ST_CurvePolygon* consist of rings that may be circularly interpolated. SQL/MM Spatial defines also corresponding WKT and WKB representations for instances of these classes. Figure 3.12 shows some examples of such geometry objects. The fifth version of SQL/MM Spatial specifies further subtypes of *ST_Curve*, e. g., elliptical curves and nonuniform rational B-splines (NURBS).

SQL/MM Spatial provides extended functionality compared with ISO 19125. Examples are

- Methods for validating geometries
- Coordinate transformations
- Approximation of circular interpolations by linear interpolations
- Support of Geography Markup Language (GML) (Chap. 4).

3-D Geometries

There is some support of two-and-a-half and three-dimensional data by current standards: SQL/MM Spatial allows

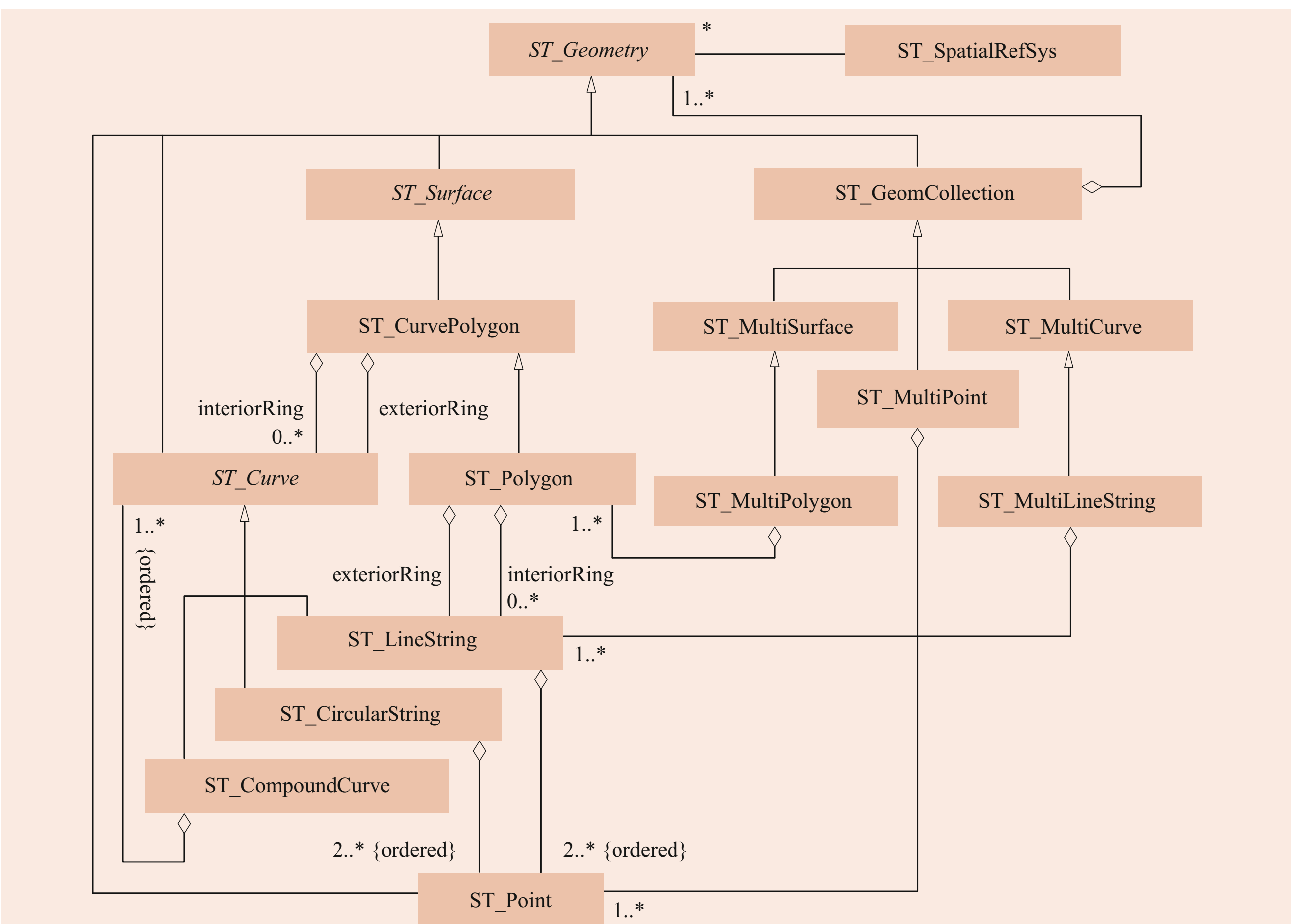


Fig. 3.11 Part of the class model of SQL/MM Spatial

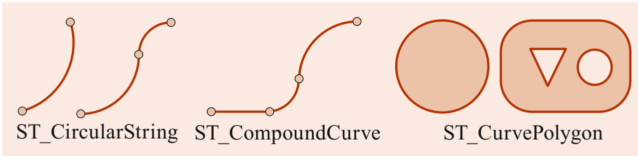


Fig. 3.12 Example of geometries with circular interpolations

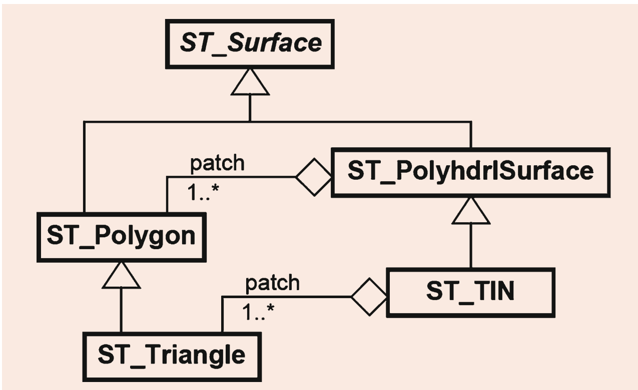


Fig. 3.13 Polyhedral surfaces and TINs in SQL/MM Spatial

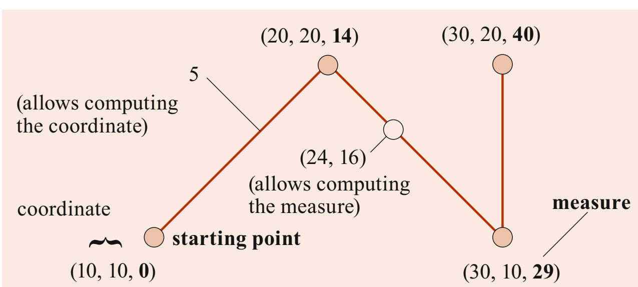


Fig. 3.14 Linear referencing system

storage of three-dimensional (3-D) points. Since the fourth version of SQL/MM Spatial, for many analysis functions, a second, three-dimensional version exists that also considers z-coordinates. Furthermore, a data model for *polyhedral surfaces* and for *triangulated irregular networks (TIN)* is part of the standard. Figure 3.13 illustrates this model: *ST_PolyhdrlSurface* consists of polygonal patches with 3-D vertices. An *ST_TIN* is a special case of a polyhedral surface; it consists of triangles instead of polygons.

ISO 19125 is purely 2-D, but the OGC has specified extensions that are similar to the data model shown in Fig. 3.13.

Reference Systems

A *spatial reference system* (SRS) allows the interpretation of stored coordinates as a description of a position in a real data space (Chap. 8). If the SRS's coordinate system is georeferenced, a coordinate describes a location with respect to the Earth's surface. Within spatial databases, *EPSG codes* are typically used for identifying spatial reference systems. These SRSs are defined by the former European Petroleum Survey Group (EPSG), today's Geodesy Subcommittee of the International Association of Oil and Gas Producers (IOGP).

In addition, a *linear referencing system* (LRS) is often provided for network data. In this case, the position of a point on a line can be derived by the position of the starting point of the line and measures that define the distance to the start (Fig. 3.14). Using this information and the geometry of the line, the coordinates of a query point can be derived (i. e., interpolated) for a given measure, or a measure can be computed for a given coordinate. SQL/MM Spatial defines the corresponding functions.

3.6.2 Topological and Network Models

The class models presented above provide methods for testing topological predicates using the geometry of the features. However, for many applications, topological data models are required. Especially the management and analysis of networks require suitable data models and operations (e. g., the computation of the shortest path; Sect. 3.9.1). Therefore, spatial database systems often offer a topological class model in addition to the basic geometry model.

SQL/MM Spatial provides the two packages *Topo-Net* and *Topo-Geo* for a network and a topological class model. Figure 3.15 illustrates how the topological primitives node, edge, and face are represented (the inheritance relationship is introduced by the author for clarity). The geometry of faces can be derived from the geometry of the corresponding edges. The relationship *containing face* allows isolated nodes to be assigned to a face. The relationship *next left edge* refers to the next edge of a face on the left (pointing from *start node* to *end node*), moving counterclockwise along the face boundary. The same holds for *next right edge* with respect to an edge on the right.

3.6.3 Raster Models

A further requirement for spatial databases is the support of raster data. Typical operations and functionality are (among others) georeferencing of raster maps, support of world and image coordinates, support of spatial queries, and decomposition and support of pyramids (Sect. 3.10.1).

Currently, no specialized data standard for raster data in spatial databases exists.

Figure 3.16 shows an overview of the data models relevant for spatial databases.

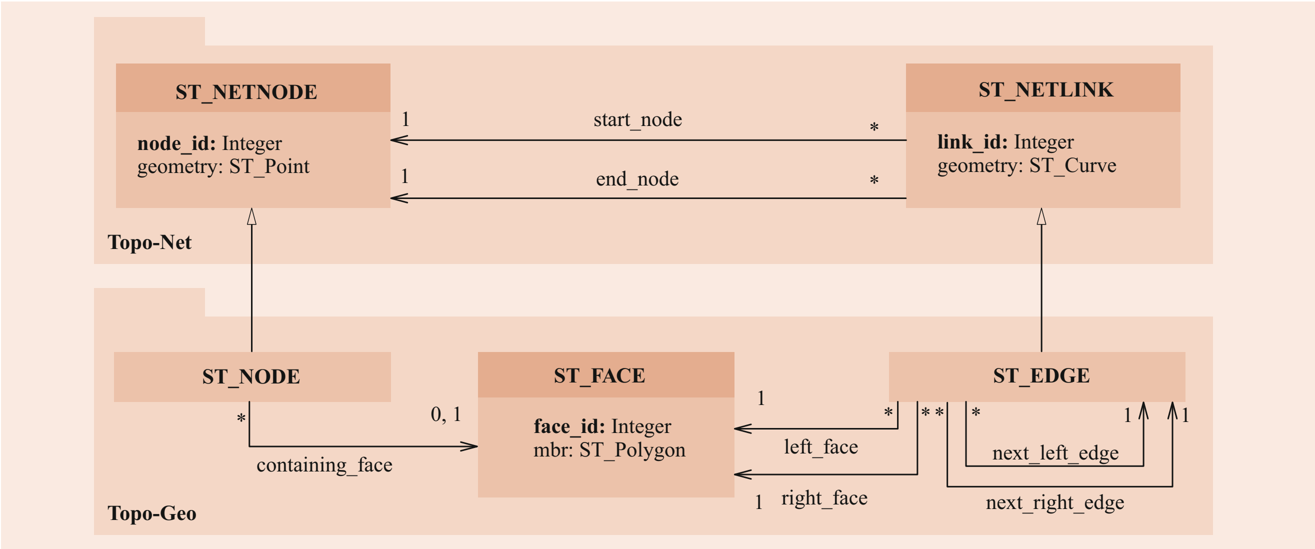


Fig. 3.15 The network and topological class model of SQL/MM Spatial

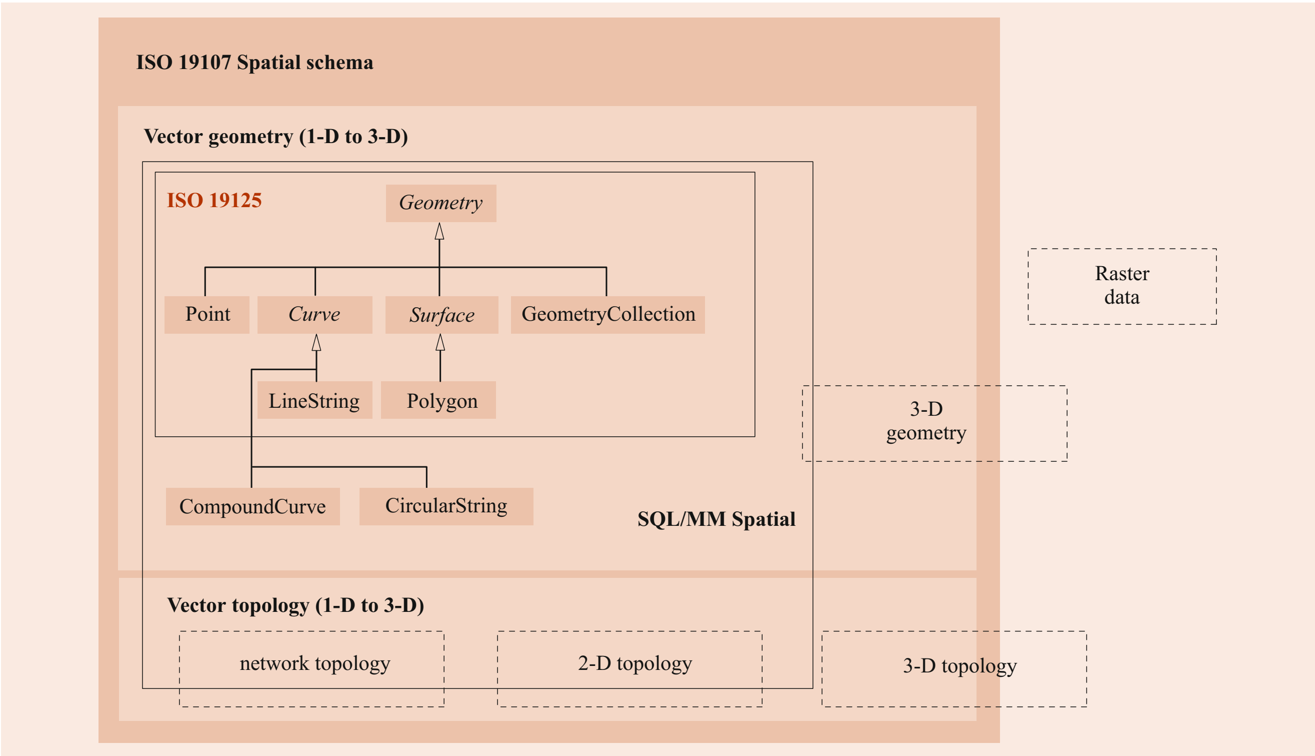


Fig. 3.16 Overview of the data models relevant for spatial databases

3.7 Spatial Query Processing

Spatial database systems enable spatial functions to be applied in SQL statements. In the case of queries, the following functions may occur:

1. In the SELECT clause
2. In the WHERE, HAVING, or/and ON clause.

In the first case, the spatial function is applied to (simple) features retrieved by the DBMS before. Therefore, the database system requires an implementation of the corresponding functionality but no further internal enhancements for processing such queries. The situation is different in the second case, because the query condition is spatial. For such a *spatial query*, the DBMS must select the relevant tuples by evaluating spatial functions. This cannot be done (efficiently) without special enhancements of the DBMS; the query has to be decomposed into one or more basic spatial queries that are processed by using special algorithms, data representations, and index structures [7, 23].

3.7.1 Basic Spatial Queries

For efficient spatial query processing, a spatial database system has to support a set of basic spatial queries.

A *point query* determines all geospatial features whose geometry contains a given query point. A typical application is the selection of areal objects by a coordinate or interactive selection. An exemplary point query in SQL notation using the geospatial features towns and highways would be

```
SELECT * FROM Towns t
WHERE Contains(t.area,
PointFromText('POINT(7 3)')) = 1;
```

A *window query* determines all features whose geometry attribute is intersected by a given query rectangle r (i. e., the geometry is completely contained by r, or it is intersected by the edges of r). Eventually, any visualization of a map by a geographic information system is based on one or more window queries. An example in SQL notation is

```
SELECT * FROM Towns t
WHERE Intersects(t.area,
PolygonFromText('POLYGON
((0 0, 0 5, 5 5, 5 0, 0 0))')) = 1;
```

The *region query* is a generalization of the window query. It determines all geospatial features whose geometry is intersected by a given query polygon. Such a polygon may be created by computing a buffer zone. A corresponding SQL statement looks like this

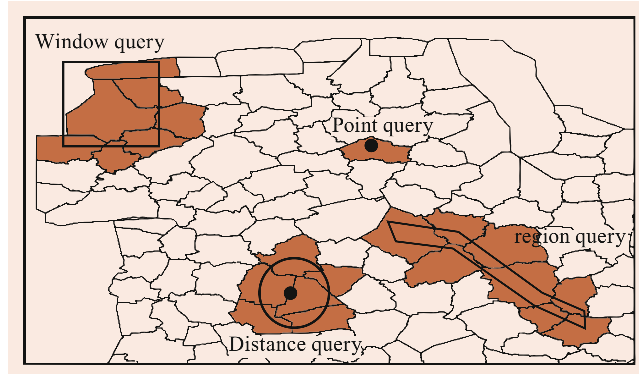


Fig. 3.17 Spatial selection queries

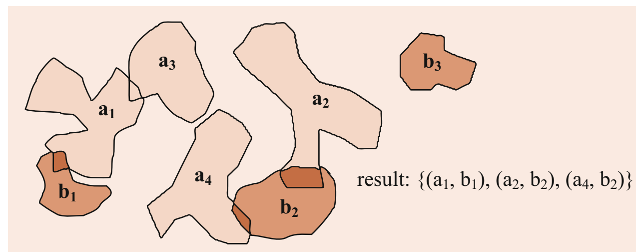


Fig. 3.18 Example of an intersection join

```
SELECT * FROM Towns t, Highways h
WHERE h.name = 'Interstate 7' AND
Intersects(t.area,
Buffer(h.geometry, 10)) = 1;
```

The *distance query* determines all features whose geometry is located within a given distance *dist* to a query point p. In other words, it is a circular region query with radius dist. The distance query can be used, for example, for the selection of lines that are near a point of interest:

```
SELECT * FROM Towns t
WHERE Distance(t.area,
PointFromText('POINT(7 3)')) < 14;
```

These four queries are *spatial selection queries*; they are illustrated in Fig. 3.17.

The *spatial join* is a join operation between two or more relations with a join condition that contains at least one spatial condition. The *intersection join* is the most important variant of a spatial join, in which the join condition tests two geometries for an intersection (Fig. 3.18). The spatial join is the base operation for performing a map overlay by a geographic information system [24]. An example in SQL notation is

```
SELECT t.*, h.*,
Intersection(t.area, h.geometry)
```

```
FROM Towns t INNER JOIN Highways h
ON Intersects(t.area, h.geometry) = 1;
```

A *nearest-neighbor query* (NNQ) determines the nearest feature or the k nearest features (k-NNQ) for a query geometry [25]. *Incremental NNQs* compute the nearest geometry first and then allow repeatedly asking for the next nearest geometry that has not yet been provided as a result object.

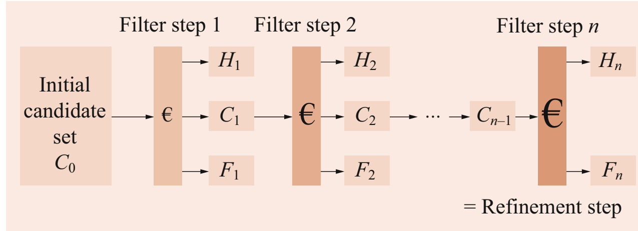


Fig. 3.19 The concept of multistep query processing

3.7.2 Multistep Query Processing

Efficient processing of spatial queries is more difficult than evaluation of typical relational query conditions. This is caused by some characteristics of geospatial data.

- The number of objects in a spatial database is often very high. In geographical applications, typically, several million geospatial objects exist.
- Compared with conventional data types of relational databases – character strings, numbers, and dates – the processing of geometrical data types is far more complex; for example, the test for the intersection of two geometries requires many times more processing time than testing whether two numbers are equal.
- Values of geometry attributes are characterized by very large variability. This observation refers to their spatial extension and to the storage requirements, as well as to their shapes and their distribution in space. Thus, we often find features in one relation whose geometries differ in their extension and/or in storage requirements by several orders of magnitude from each other. Frequently, areas exist in which a high concentration of geospatial features exists, for example, in the downtown area of a large city.
- As the capture of geographical data is very expensive, geospatial datasets are usually used on a long-term basis. Nevertheless, the features (and their geometry) are subject to updates to correct errors or to adapt to changes of the modeled world.

These characteristics must be considered when designing a spatial database system. One consequence is the multistep approach of spatial queries.

Concept of Multistep Query Processing

Multistep query processing means that a query is processed by a sequence called *filter steps*. The general objective is to limit the number of objects that potentially fulfill the query condition (called *candidates*) in early filter steps through the use of fast algorithms and suitable data representations. Candidates that are later recognized as not fulfilling the query condition are called *false hits*. An additional goal of this strategy is the identification of as many hits as possible by the filter steps; a *hit* is an object that reliably fulfills the query. In the last step, called the *refinement step*, the exact and expensive test of the query condition is finally done on a (it is hoped, significantly) reduced candidate set [26].

The concept of multistep query processing is illustrated schematically in Fig. 3.19. The query is processed in n filter steps; C_i, H_i, and F_i describe the sets of candidates, hits, and false hits after the i-th filter step. The combination of the hits H_i $(1 \leq i \leq n)$ forms the complete result of a query. It holds that

$$C_i = H_{i+1} \cup F_{i+1} \cup C_{i+1} (0 \leq i < n).$$

In Fig. 3.19, the size of the Euro symbol illustrates the cost required for processing a candidate in a filter step. This [central processing unit (CPU) and input/output (I/O)] cost increases through the sequence of filter steps. Therefore, it is an important goal to reduce the total cost for a query by identifying a large number of hits or false hits in early filter steps.

Typical filter steps of a spatial database system are (in this order) the use of

- A spatial index (Sect. 3.8)
- Simple and effective approximations of the geometry (Sect. 3.7.3)
- Efficient and robust algorithms from the field of computational geometry.

3.7.3 Filtering by Approximations

As was mentioned before, an essential difficulty of spatial query processing is caused by the complexity of geometries. Therefore, it is a common approach in spatial databases to store an *approximation* that describes the exact geometry by a simplified geometry. The approximation defines roughly the position as well as the extension of the original object [27].

The most common approximation is the *minimum bounding rectangle* (MBR). The border of the MBR is aligned with the axes of the coordinate system. Figure 3.20a shows an approximation to Great Britain by an MBR.

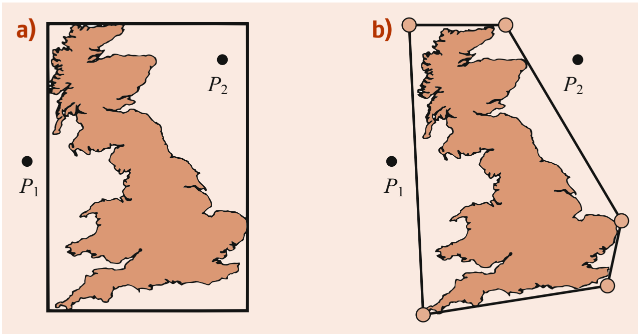


Fig. 3.20 Minimum bounding rectangle **(a)** and minimum bounding pentagon **(b)**

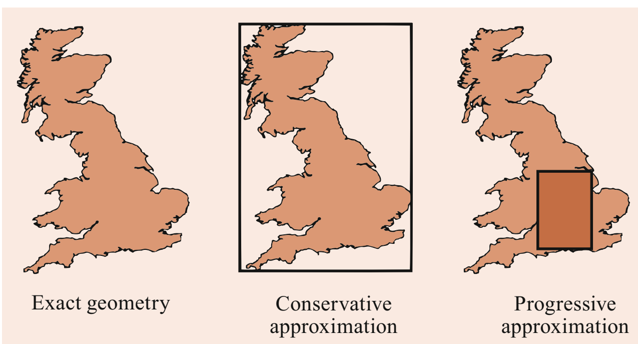


Fig. 3.21 Conservative and progressive approximation

In the context of a multistep query, a filter step tries to identify as many features as possible on the basis of their approximation as hits or as false hits. Computation using the approximation is typically far less expensive than the comparable operation using the exact geometry. An MBR often allows it to be determined whether the geometry of a feature fulfills or fails the query condition. However, an approximation does not always identify an object as a hit or a false hit. Then, the feature remains in the candidate set for the next filter step.

The more accurate the approximation, the better the filter step works. In Fig. 3.20, the MBR allows the depicted geometry to be identified as a false hit for query point P_1 but not for query point P_2. In contrast, the more precise pentagon enables the query processor to identify both query points as false hits. Thus, a tradeoff exists between the quality of an approximation on the one hand and the cost of its computation, its evaluation, and its storage requirements on the other hand. Regarding the minimum bounding pentagon, for example, it is more expensive to compute it from the exact geometry, more costly to test it for spatial query conditions (such as the test for intersection), and it requires (by a factor of 2.5) more storage space than the MBR. The space requirements are especially important for efficient use within a spatial index (Sect. 3.8).

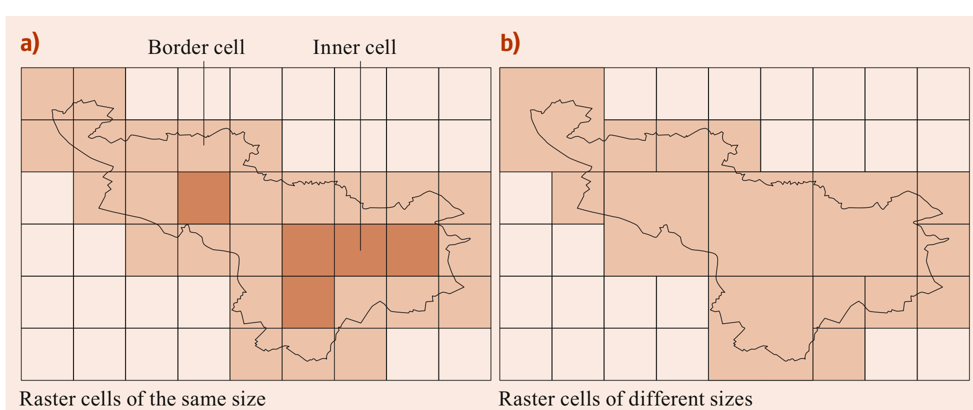
Conservative and Progressive Approximations

The minimum bounding rectangle is a *conservative approximation*, i. e., the approximated geometry is completely contained in the approximation. Conservative approximations primarily allow exclusion of false hits from the candidate set. A *progressive approximation* is completely included by the original geometry. Contrary to conservative approximations, progressive approximations often enable identification of hits; for example, if the query point of a point query is within a progressive approximation, then it is also included by the exact geometry. The intersection of two progressive approximations implies that associated geometries also intersect. Figure 3.21 depicts examples of conservative and progressive approximation.

Simple and Composed Approximations

A further relevant distinguishing factor between approximations is the number of components of which they consist. *Simple approximations* consist of a single component (such as all approximations presented so far in this section). *Composed approximations* describe geometries by a collection of elementary geometries, such as triangles or rectangles. Often, composed approximations are based on the raster model. Then, raster cells or similar regions are used as components. Composed approximations may only consist of components

Fig. 3.22 Composed approximations



of the same size (and form) or of components of different sizes (and forms). By including and tagging components that intersect the border of the original geometry, composed approximations allow conservative and progressive approximation to be used at the same time. Figure 3.22a depicts an approximation composing raster cells of the same size and distinguished in *border cell* and *inner cell*. Figure 3.22b shows a composed approximation with raster cells of different sizes.

The higher the number of components, the smaller the elementary components and the better the quality of the approximation. Again, we can observe a trade-off between quality on the one hand and cost for computation, evaluation, and storage requirements on the other. In the case of composed approximations, it must be taken into account that more than one component may allow the object to be identified as a hit. This redundancy requires the elimination of duplicates from the result set.

3.8 Spatial Indexing

For efficient processing of spatial queries, a spatial database system must be able to index the spatial attributes that are used in the filter conditions of queries. As introduced in Sect. 3.2.8, relational database systems typically use search trees, such as B-trees or their variants as index structures. For spatial queries, however, pure search trees are not adequate for indexing, because search trees use a linear order for sorting and organizing data. For relational data types such as numbers and character strings, a linear ordering is obvious, but for geometric data types this is not the case; for example, it is not clear whether the location of the Eiffel Tower is smaller or larger than that of Tower Bridge. Such a comparison is even more difficult to define for areas. Therefore, we need special techniques and data structures for efficient spatial indexing [7].

Suitable *spatial index structures* should observe the following requirements.

- Points and approximations of curves and areas have to be organized.
- Basic spatial queries (Sect. 3.7.1) must be efficiently supported. Such queries typically require access to spatially close data. Therefore, a direct consequence is that proximity should be maintained, i. e., spatial features in the same neighborhood should be stored with high probability in the same node of an index structure.
- Dynamic inserts, modifications of spatial objects, and delete operations need to be supported. The efficiency of the index structure should not be significantly influenced by the number, type, and order of update operations.
- Reasonable storage utilization should be guaranteed.
- The index structure should be robust to nonuniform data distributions; the storage requirements and the processing time of queries should not significantly increase in such cases.

The entries in a node of an index structure are stored in a block in secondary storage. Thus, a node corresponds to a database block, which is also called a bucket. In the case of a spatial index structure, each node is associated with some region of the data space. This region is called a *block region* or *bucket region*. In the case of a tree, the block region of a directory node encloses all block regions of the nodes of its subtree.

3.8.1 Basic Techniques

For the organization of spatial data by index structures, several basic techniques have been proposed in the literature [28].

Clipping Technique

The *clipping technique* assigns any geometry (or its approximation) to each block region that it intersects. Figure 3.23 outlines this idea. In the example, the rectangle R_1 is assigned to four block regions. The main disadvantages of the clipping technique are that the number of entries increases faster than the number of stored objects and that duplicates may be generated.

Point Transformation

The point transformation converts a d-dimensional rectangle (e. g., the MBR of a geometry) into a two-dimensional point, i. e., it transforms it into a higher-dimensional space. We can distinguish two variants.

- The *corner transformation* represents a rectangle by its opposing corner points; in the case of a two-dimensional rectangle by the points $(x_{\min}, y_{\min})$ and $(x_{\max}, y_{\max})$.
- The *center transformation* describes a rectangle by its center point and the half-extensions in each dimension, in 2-D denoted by $(x_{\text{cent}}, y_{\text{cent}})$ and $(x_{\text{ext}}, y_{\text{ext}})$.

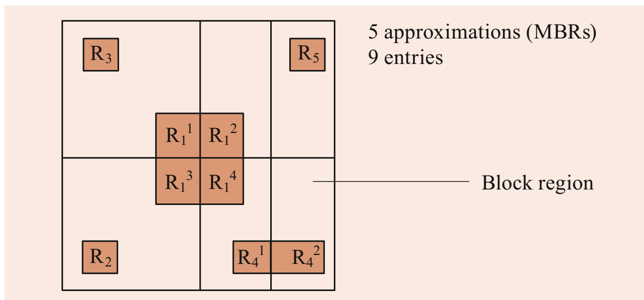


Fig. 3.23 Example of the clipping technique

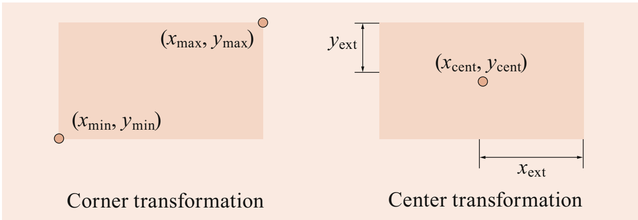


Fig. 3.24 Example of the corner **(a)** and the center transformation **(b)**

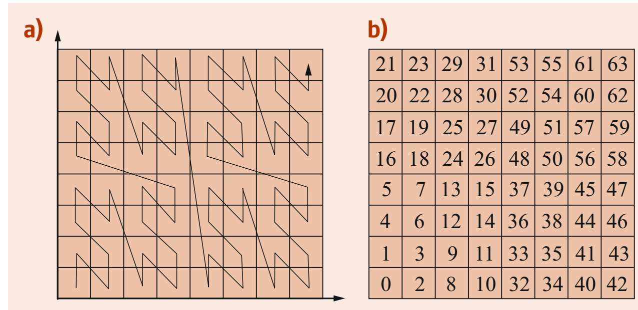


Fig. 3.26 The z-order curve **(a)** for deriving cell indices and an ordering of the cells **(b)**

Figure 3.24 illustrates both variants for a two-dimensional rectangle.

The resulting points can be stored by an index structure for multidimensional points. The main disadvantage of this approach is that proximity may be lost by considering the extension of the rectangles. Figure 3.25 demonstrates this effect for one-dimensional rectangles (i.e., for intervals). Originally, I_1 intersects I_2 and I_3, but I_2 does not intersect I_3. In the transformed data space, however, the corresponding points P_2 and P_3 are closer to one another than they are to P_1.

Embedding into One-Dimensional Space

Space-filling curves (also called *Peano curves*) allow linear ordering of multidimensional points. Examples are curves based on the *z-order* (also called *Morton order*) and the *Hilbert curve* (Sects. 3.8.2 and 3.8.4).

To apply space-filling curves, a d-dimensional data space is divided into 2^r rectilinear cells; r defines the *resolution* of the curve. The space-filling curve passes through each of the cells once, which allows a *cell code* and a linear ordering of the cells to be derived. Figure 3.26 illustrates this approach for the z-order with resolution 6.

The cell code can be assigned to each of the geometries contained by the cell. Thus, the geometries can be sorted and organized by a traditional search tree. However, there are some problems and challenges with using space-filling curves for spatial indexing. First, spatial proximity is not maintained in all areas. In Fig. 3.26, for the quadratic region covered by cells 4–7 the cell indices are close, but for the quadratic region defined by cells 15, 26, 37, and 48, this observation does not hold.

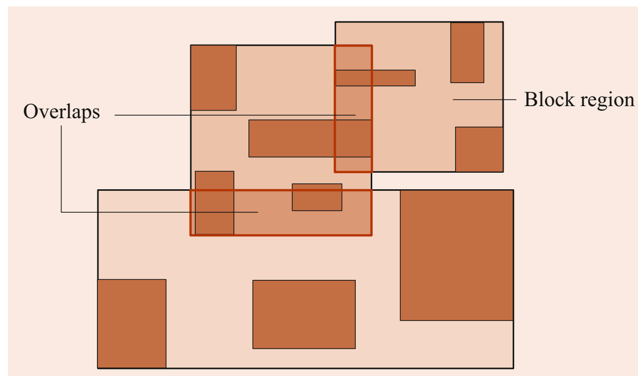


Fig. 3.27 Example of the overlapping-regions approach

The second challenge is the handling of shapes that have an extension and may be intersected by more than one cell. Third, a uniform partition of a data space does not support nonuniform data distributions well.

Overlapping Regions

The basic idea of this technique is to allow *overlapping block regions*, so that the partitioning of the data space is no longer disjoint. Such overlap avoids arbitrary cutting of extended shapes. Figure 3.27 depicts an example.

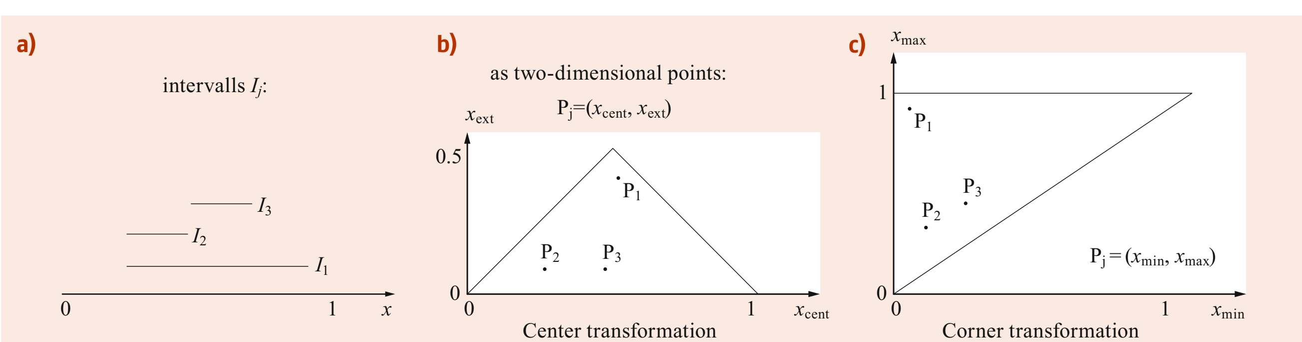


Fig. 3.25 Loss of proximity in the transformed data space

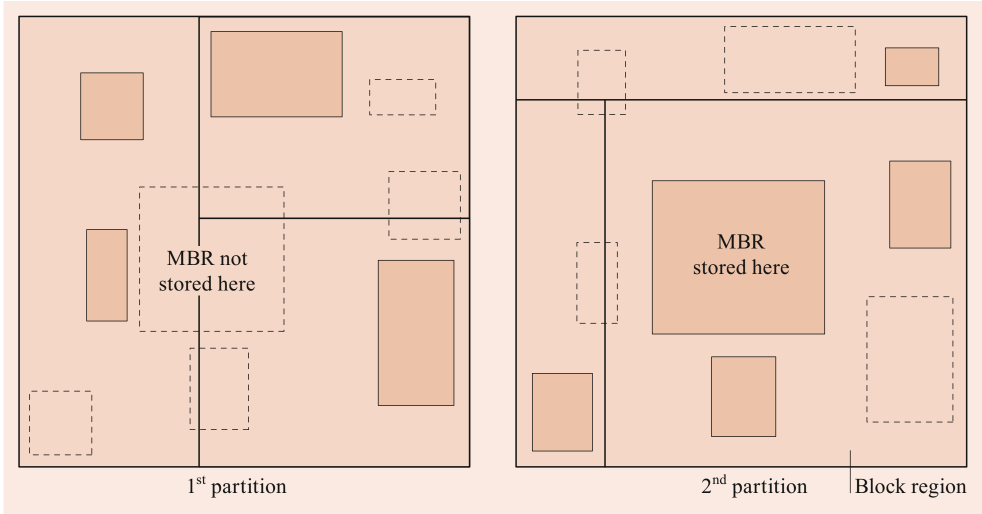


Fig. 3.28 Example of the multipartition technique

The main disadvantage of this approach is that a query covering an overlap has to process all elements of the index structures involved in the overlap. Therefore, from the point of view of performance, it is essential to minimize overlaps.

Multipartition Technique

The *multipartition technique* avoids cutting extended geometries by organizing several space partitions in parallel. A spatial object will be inserted into a partition with a block region that includes its geometry completely. The block regions are free of overlaps in each individual partition. All of them organize the same data space, i. e., we have an overlap between them. Figure 3.28 illustrates the approach using an example with two partitions.

Because queries have to search in all partitions, it is important to keep their number extremely small. However, the smaller the number of partitions, the more difficult it is to find partitions that do not cut any of the geometries.

Comparison

Finally, Table 3.16 compares the five presented basic techniques for spatial indexing with respect to duplication of geometries, preservation of proximity, and overlap ("+" indicating a positive characteristic).

Table 3.16 Evaluation of the basic techniques for spatial indexing

Technique	No duplicates	Proximity	No overlaps
Clipping technique	−	+	+
Point transformation	+	−	+
Embedding	−	±	+
Overlapping regions	+	+	−
Multipartition technique	+	+	−

3.8.2 Quadtrees

Quadtrees are spatial data structures that divide a d-dimensional data space recursively into 2^d cells of equal size. In the three-dimensional case, they are also called *octrees*. Quadtrees are used for many applications in the fields of computer graphics and image processing. There are several variants of quadtrees to organize points, curves, and areas [29].

Figure 3.29 shows a point region (PR) quadtree storing point geometries. The two-dimensional PR quadtree divides a cell recursively into four quadrants until no cell contains more than one point. The cells of the next level are marked according to their respective direction with NW, NE, SW,

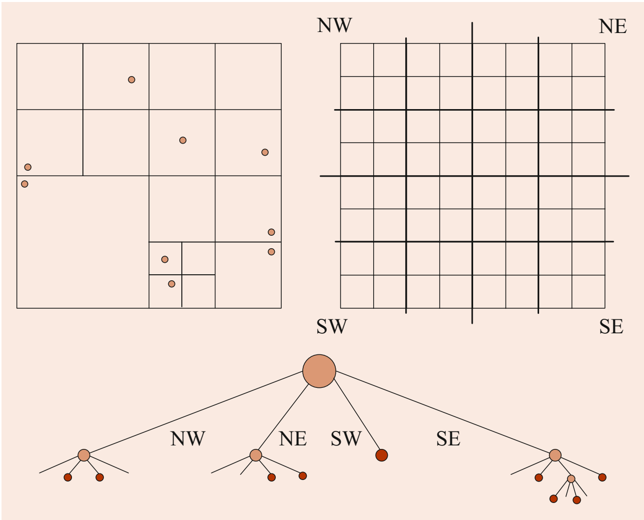


Fig. 3.29 Example of a point region (PR) quadtree

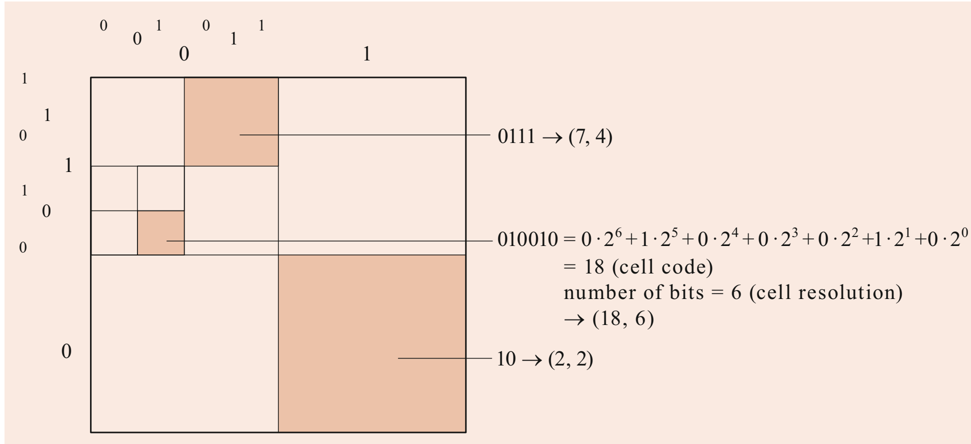


Fig. 3.30 Computation of z-values

and SE. This procedure is represented by a corresponding tree: The root node with its four child pointers symbolizes the partitioning of the complete data space into four quadrants. The leaf nodes of this tree store a point (dark circle) or are empty, while the inner nodes (brown-shaded circle) represent a partitioning of a cell into four quadrants.

Linear Quadtrees

To use a quadtree as an index structure for spatial databases, the tree must be able to store several geometries per node. This can be achieved by linear quadtrees. *Linear quadtrees* map the data structure of a quadtree into a one-dimensional data space and organize the result by a one-dimensional index structure (mostly by a B$^+$-tree).

Space-Driven Linear Quadtrees

Space-driven linear quadtrees describe the quadtree cells by cell indices computed using the z-order and organize them by the aforementioned one-dimensional index structure.

Because of the fact that in a quadtree, cells of different sizes may occur, a cell must be described not only by the cell code but also by its size using the resolution of the cell. The pair (cell code, cell resolution) is called the *z-value*. The cell code can be easily computed by repeatedly bisecting the data space until the cell is reached. By calculating for each bisection whether the cell is located to the left of, under the current partition line ($= 0$), to the right of, or above the current partition line ($= 1$), a bit sequence is computed. With the last determined bit having the lowest value ($= 2^0$), the interpretation of this bit sequence as an integer number provides the cell code. Its length corresponds to the cell resolution. Figure 3.30 illustrates this process.

A comparison of z-values is required for z-ordering. To compare z-values of different resolutions, the cell code with the higher resolution is divided by $2^{\Delta r}$, where Δr is the positive difference between the two resolution values. In other words, the lower bits of the longer bit sequence are cut in order to obtain sequences of the same length.

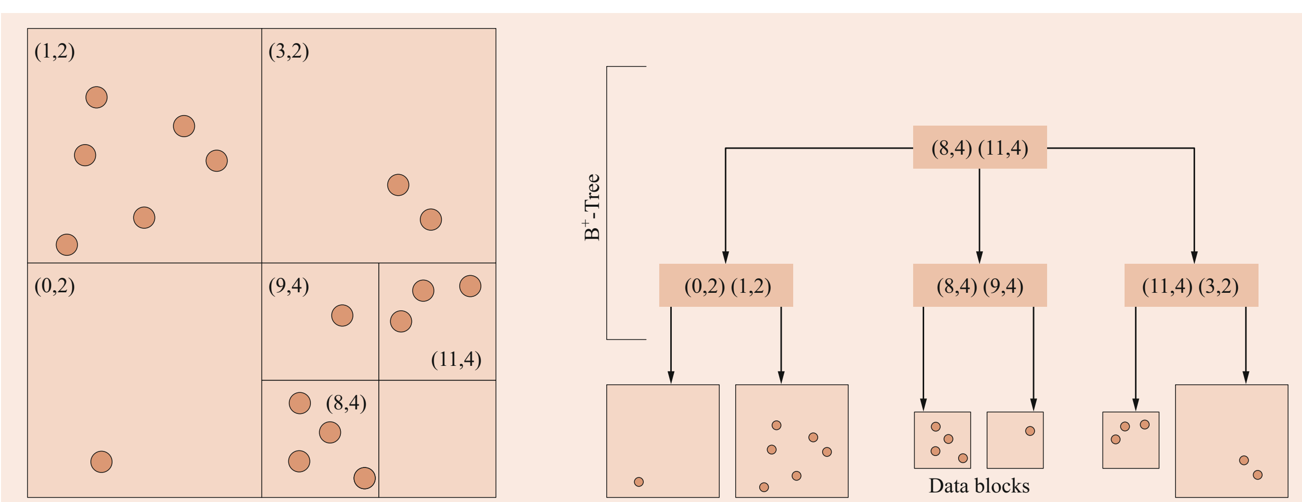


Fig. 3.31 Example of a space-driven linear quadtree

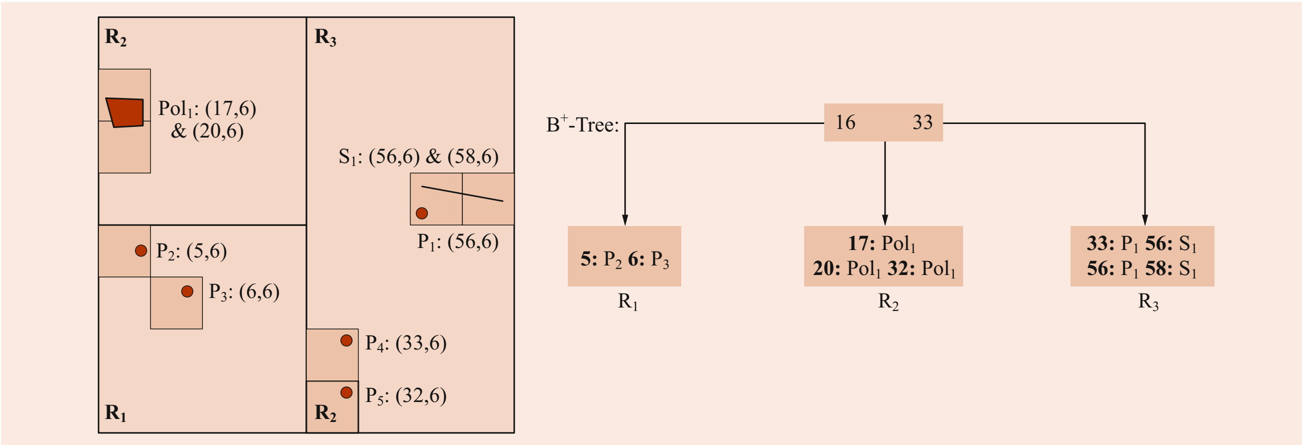


Fig. 3.32 Example of a data-driven linear quadtree

Figure 3.31 depicts an example where, for each leaf node with data, the z-value is computed and organized by a B$^+$-tree. The example shows that this approach may lead to data blocks with strongly varying storage utilization.

Data-Driven Linear Quadtrees

Data-driven linear quadtrees approximate the exact geometries by one or more quadtree cells, i. e., they compute composed approximations (Sect. 3.7.3). These cells are described by z-values and organized by a one-dimensional index structure. Figure 3.32 illustrates this idea by using z-values of the same resolution. The example emphasizes that nonrectangular block regions (R_3) and noncontinuous block regions (R_2) may occur. By using several cells for describing one geometry and not storing them together, a spatial query processor may determine duplicates that must be eliminated.

A data-driven linear quadtree can also be implemented by using z-values of different resolutions.

3.8.3 Grid Files

The *grid file* is an index structure for the storage of d-dimensional points. It consists of a *grid directory*, which holds references to data blocks. The grid directory subdivides the data space by a d-dimensional grid. *Linear scales* define the subdivision of the grid in each dimension [30]. These terms are explained in Fig. 3.33.

Each grid cell of the grid directory stores exactly one reference to a data block. In order to avoid low storage utilization of the data blocks, more than one grid cell may refer to the same data block. The cells pointing to the same data block form a block region. The block regions are always d-dimensional rectangles. In Fig. 3.33, block regions are represented by rounded-off rectangles.

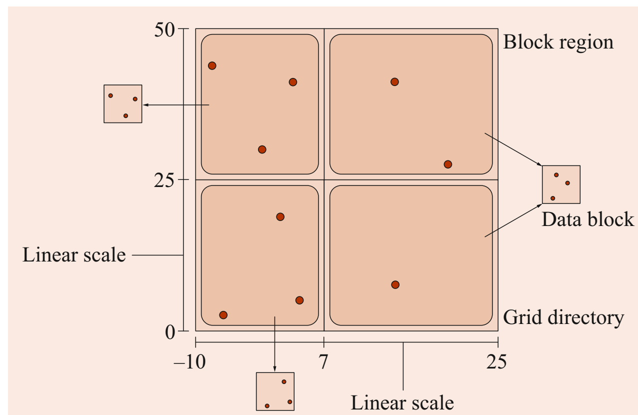


Fig. 3.33 Structure of a grid file

Processing of Spatial Queries

Processing of queries is relatively simple. For point queries, the grid cell containing the query point is determined by using the linear scales. Then the directory block in which the corresponding grid cell is stored is read from secondary storage. Using the reference stored in the grid cell, the actual data block can be accessed. With the linear scales cached in the main memory, two disk accesses are necessary. Window queries may require a larger number of grid cells and the retrieval of all referenced data blocks. Consequently, more disk accesses are necessary.

Insert Operation

When inserting a new point, the appropriate data block is determined with the help of the linear scales and the grid directory. The insertion is completed if the data block has sufficient space to store the new point. Otherwise, an overflow of the data block occurs, so that it must be divided into two blocks. Two cases may occur.

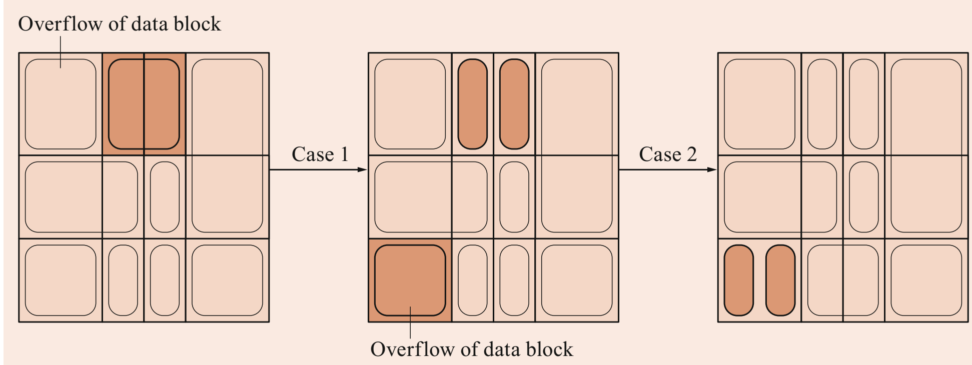


Fig. 3.34 Two types of overflow handling by grid files

1. The block region of the overflowing data block includes more than one grid cell. In this case, a partitioning line is activated, which cuts the block region and divides the data block into two blocks, as shown for the left overflow in Fig. 3.34. At least one data point is required in each of the new block regions. Data blocks with low storage utilization may be generated.
2. The block region of the overflowing data block includes only one grid cell. Now, a new partitioning line must be introduced. This split results in a new $(d-1)$-dimensional row or column, into which (apart from the overflowing grid cell) the references of the split row or column are duplicated. Figure 3.34 illustrates this procedure by the right overflow. There is some degree of freedom in determining the new partitioning line. The choice of a partitioning line consists of selecting the split axis and computing the split position. Especially in the case of nonuniform data distributions, many grid cells storing duplicate references may be created.

Different techniques for organizing the grid directory as a dynamic multidimensional array have been proposed in the literature. Many authors recommended using a multilevel grid directory in order to restrict the effect of a new partition line to a small area of the grid directory. Then, the grid file becomes less sensitive to nonuniform data distributions.

3.8.4 R-trees

R-trees organize d-dimensional rectangles by allowing overlapping block regions. An R-tree is a balanced tree whose leaf nodes all have the same distance to the root node [31, 32]. As for B$^+$-trees, data nodes and directory nodes are distinguished.

- A *directory node* stores entries e that consist of the reference (e.ref) to the direct child node and a block region described by the minimum bounding rectangle (e.mbr) of all entries in the child node and, so, of the corresponding subtree.
- A *data node* contains entries that store a minimum bounding rectangle as an approximation of the actual spatial feature. Furthermore, an entry contains the spatial feature or (more often) a reference to it.

The root node stores at least two entries unless it is the only data node. Otherwise, each node has between $\text{cap}_{\min}$ and $\text{cap}_{\max}$ entries with $\text{cap}_{\min} \leq \text{round}(\text{cap}_{\max}/2)$. Like B-trees, R-trees guarantee storage utilization of at least $\text{cap}_{\min}/\text{cap}_{\max}$ and a height growing logarithmically with the number of stored objects. R-trees are completely dynamic: insert and delete operations can be mixed with any kind of spatial query.

Figure 3.35 illustrates an R-tree of height 3, which consists of three directory and four data nodes. In the upper part of Fig. 3.35, the partitioning of the data space on the different tree levels is shown. The spatial features' minimum bounding rectangles are shaded in brown. The block regions are shown as dashed lines. The lower half of Fig. 3.35 depicts the corresponding data structure.

Processing of Spatial Queries

The processing of window and point queries starts at the root node and determines all entries intersected by the query window or containing the query point. The nodes referenced by these entries are successively read into the main memory and processed like the root. The process is continued recursively until the data nodes are reached. Entries of the data nodes are included as candidates in the result set for further processing (Sect. 3.7.2). Because of the overlapping block regions, a point query may branch into more than one subtree.

Insert Operation

An insert operation has to determine the data node that accommodates the new object. If the MBR of an entry on the path from the root node to the destination node does not contain the MBR of the new object completely, the entry's MBR will be correspondingly enlarged. If the insertion of the new object leads to an overflow of the destination node, i. e., if its maxi-

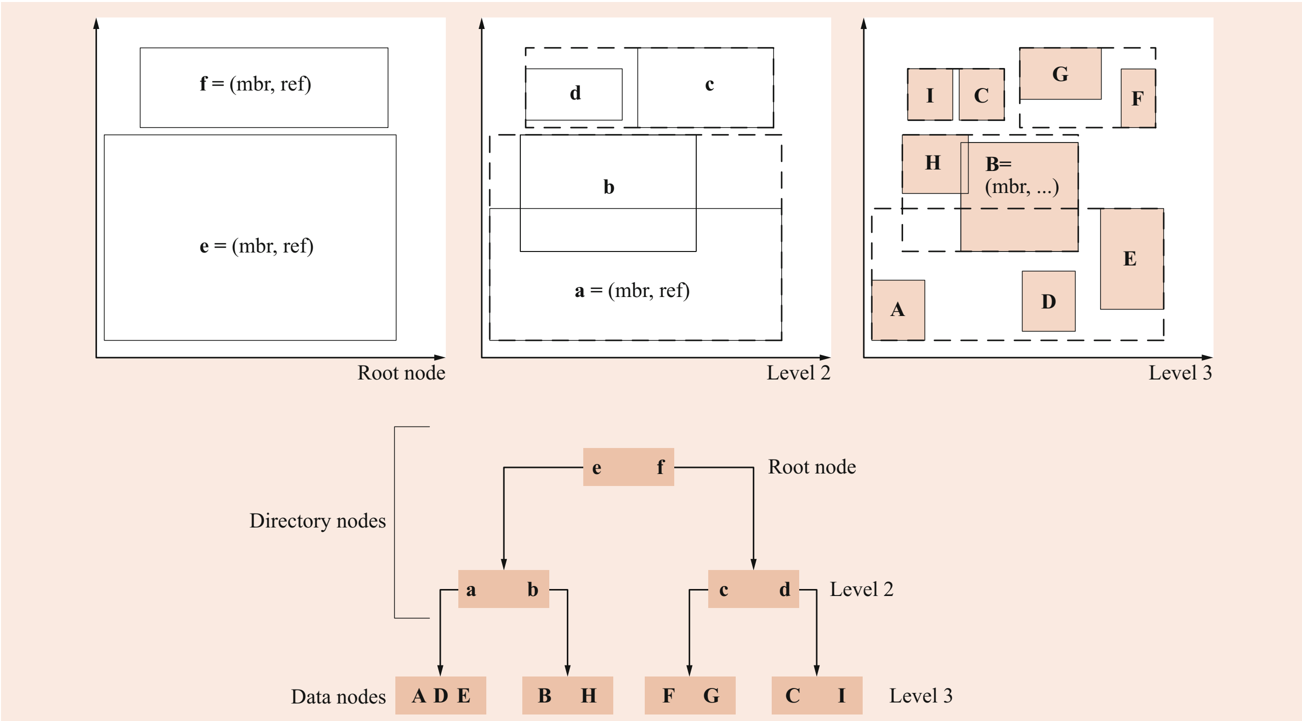


Fig. 3.35 An example of the data structure of an R-tree

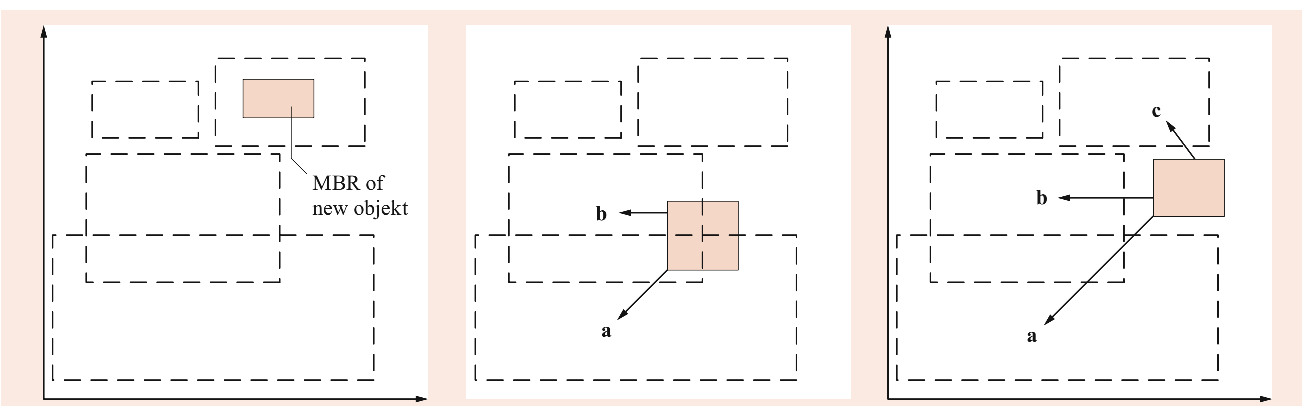


Fig. 3.36 Different cases for the choose-subtree operation

mum capacity cap_{max} is exceeded, the node will be split into two nodes, and a corresponding new entry will be stored in the parent node. Like in B^+-trees, this may lead to an overflow in the parent node. Nevertheless, the processing is restricted to a single path from the root to the destination node.

The design of two operations is essential for the performance of an R-tree.

1. The *choose-subtree operation* determines the data node or the subtree (i. e., the directory entry) that should accommodate the new object. Three cases may occur: the object MBR is contained in one or several entry MBRs or in none. For the two latter cases, some degree of freedom for the entry selection exists (Fig. 3.36).
2. In the case of an overflow, the *split operation* has to compute which of the entries moves into the new node.

There are different variants of R-trees in the algorithms used for the choose-subtree and the split operations.

R^*-Tree

The R^*-tree is a popular and very query-efficient variant of the R-tree [33]. Its choose-subtree and split operations try to consider four optimization criteria.

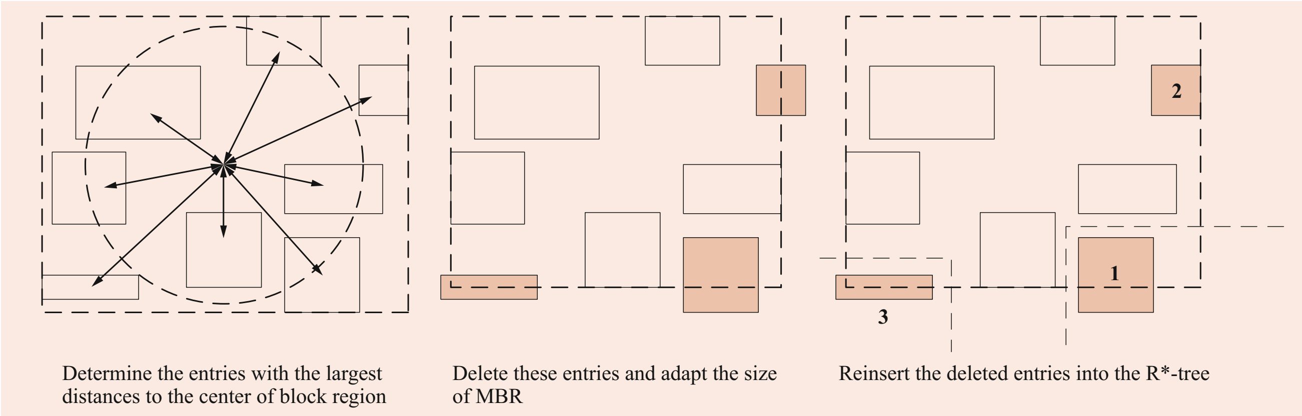


Fig. 3.37 Processing of the forced reinsert operation

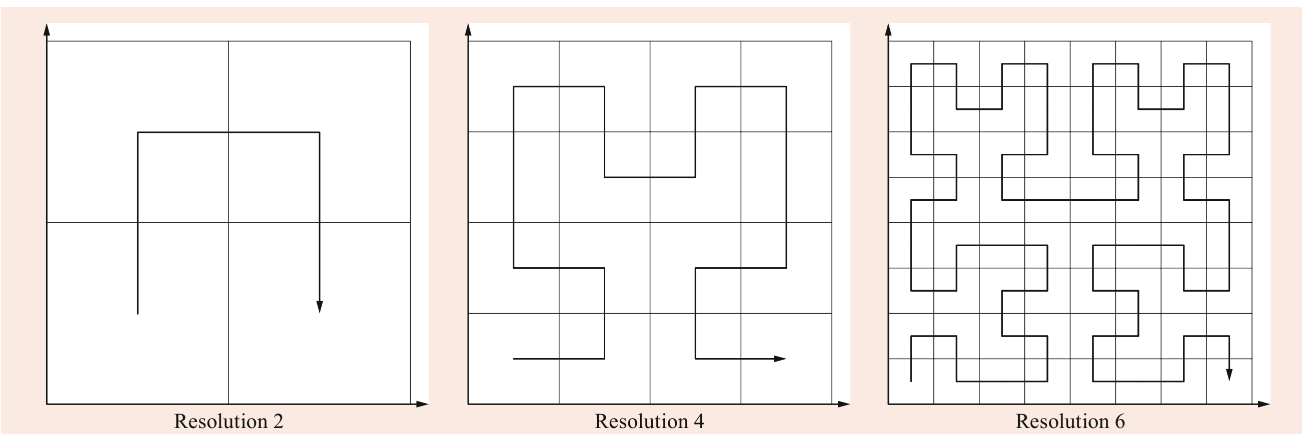


Fig. 3.38 Refinement of the Hilbert curve

1. Minimization of the overlap between block regions. A minimized overlap reduces the probability of branching into several paths during spatial queries.
2. Minimization of the area of block regions. The smaller a block region, the more rarely its subtree will be accessed during query processing.
3. Minimization of the perimeter of block regions, i. e., the block regions should become quadratic. This is motivated by query windows that are often approximately quadratic themselves. In this case, the query window intersects (on average) fewer block regions.
4. Maximization of the storage utilization. High storage utilization also reduces the number of blocks accessed during spatial queries.

None of the four criteria leads to good performance by itself. Only optimizing the storage utilization, e. g., may result in very poor query efficiency. Therefore, the R^*-tree tries to find a balance between the four criteria.

The R^*-tree introduces a further concept. The *forced reinsert* reorganizes the tree step-by-step by deleting and reinserting entries. This allows the tree to adapt to modified data distributions (maybe resulting from spatially sorted inserts). This operation is called when an overflow occurs before the split is performed. The entries with the largest distances to the centroid of the block region are deleted, and the MBRs of the corresponding parent entries are adapted. Then, the deleted entries are reinserted into the tree. Note that the reinsert operation stops at the original level of the entries and that it cannot trigger a further forced reinsert on the same level (but on upper levels). In many cases, the reinserted entries will be stored in nodes other than the original node, and a split operation can be avoided. An important disadvantage of the forced reinsert is that the processing of an insert will not be restricted to one path from the root to the destination node. Figure 3.37 illustrates the forced reinsert; in this example, entry 2 is stored in the same node as before, whereas entries 1 and 3 are inserted into other nodes.

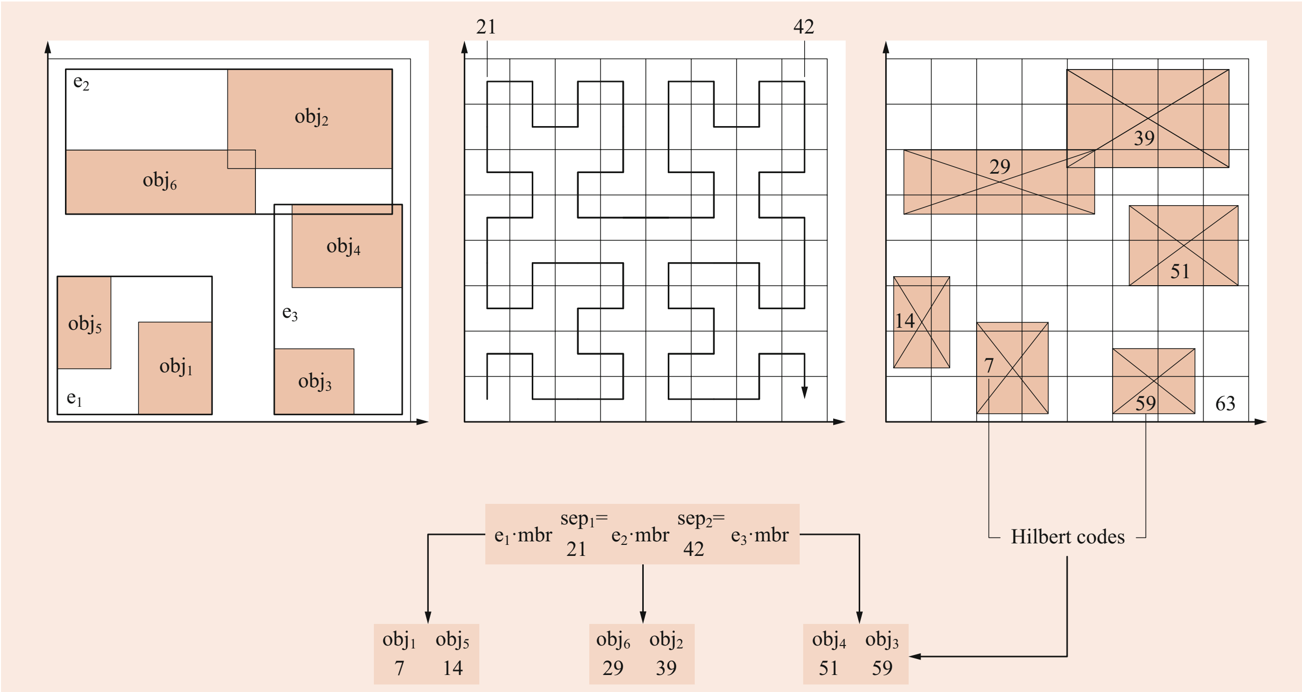


Fig. 3.39 Example of a Hilbert R-tree

Hilbert R-Tree

The *Hilbert R-tree* is a mixture of an R-tree and a B-tree [34]. It exploits the *Hilbert curve* as a space-filling curve. The Hilbert curve preserves proximity (slightly) better than *z*-ordering, but it is more complicated to calculate. Figure 3.38 depicts the Hilbert curve for three different resolutions.

The entries in the directory nodes of the Hilbert R-tree consist of a reference to the child nodes (like in R and B-trees), of a minimum bounding rectangle (as in the R-tree), and of a separator (like in the B-tree). For all MBRs in a data node, the Hilbert code is determined for the centroid of the MBR. The *Hilbert code* is the cell code determined by using the Hilbert curve. The separator works for the Hilbert codes like the separator in a B^+-tree. Figure 3.39 shows an example of a Hilbert R-tree.

The insertion of a new object and overflow handling is performed like in a B^+-tree, using the Hilbert code of the new object and the separators stored in the directory nodes. In addition, the MBRs in the directory nodes are readapted if necessary. Spatial queries are performed like in R-trees using the MBRs.

The main advantages of the Hilbert R-tree are its simple implementation based on a B^+-tree and its fast insert algorithm. However, the overlap between the MBRs is typically higher than in R-tree variants using more sophisticated choose-subtree and split algorithms. Therefore, the Hilbert R-tree has inferior search performance compared with the R-tree.

3.9 Network Databases

Spatial network databases combine the representation of the network topology with geospatial information such as position and extension. Such databases can serve, for example, storage and analysis of road networks or telecommunication networks. Traditionally the main application of *network information systems* (NIS) is the capture, documentation, management, and analysis of utility networks [7, 23].

As introduced in Sect. 3.6.2, with *Topo-Net* SQL/MM Spatial provides a class model that allows storage of a network as a *graph* that consists of *nodes* and *edges*. As basic operations, a network model allows the graph to be passed by determining the edges and nodes adjacent to a node and an edge, respectively. A *path* from a starting node *s* to a destination node *d* is a sequence of neighboring nodes, with *s* as the first and *d* as the last node in this sequence.

For many network applications, it is necessary to consider the direction of edges. Then, the corresponding graph is called a *directed graph*. Often a (static or dynamic) representation of the cost of passing edges and nodes are required. Such a graph is called *weighted*. The *path length* sums up the weights of all nodes of a path and of all edges between two neighboring nodes of the path. Figure 3.40 illustrates the terms *path* and *path length*.

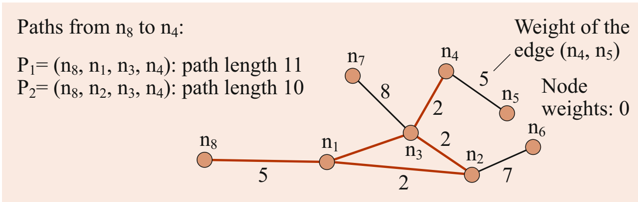


Fig. 3.40 Paths and path lengths

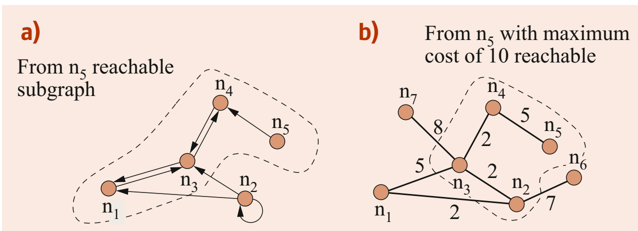


Fig. 3.41 Accessibility in networks

Fig. 3.43 Traveling-salesman problem for six nodes

3.9.1 Network Analysis

On top of a network data model, more sophisticated network analysis functions can be implemented by a spatial network database system:

The accessibility of a node d from a node s corresponds to the question of whether there is (at least) one path from s to d. The determination of all nodes that are reachable from a starting node s is a related question; in other words, the largest *connected subgraph* that includes s is computed (Fig. 3.41a with n_5 as starting node). Such an operation can be used for controlling the quality of network data. A variant of this analysis function determines all nodes or edges that are reachable from node s within a given maximum path length (Fig. 3.41b). In the context of location-based services, the determination of all points of interest that can be reached from the current position within a given time is an example of an application of this operation. It also allows the determination of zones of attraction for location studies.

One of the most important network analysis operations is the computation of the *shortest path* between two nodes, i. e., the path with the lowest cost. In Fig. 3.40, P_2 is the shortest path between n_8 and n_4. The computation of shortest paths is the major task of every navigation system. Then, the expected speed on a road determines the weight of an edge. The computation of all paths between two nodes remaining below a given cost limit is an interesting variant of the shortest-path problem.

The *spanning tree problem* in an undirected connected network asks for a tree connecting all nodes of the network. The nodes and edges of the tree are a subset of the complete network. Figure 3.42 shows two possible spanning trees for an exemplary network. The most important variant of this operation is the computation of the *minimum spanning tree*, which is the spanning tree with the minimum sum of edge weights. Such trees, e. g., are needed for systematic planning of communications networks.

The *traveling-salesman problem* is to compute the shortest path within a network that visits all given destination nodes and returns to the starting node (Fig. 3.43). Due to the exponential complexity of exact algorithms for this problem, often, heuristic solutions that result with a high probability in an almost optimal solution are used. Similar tasks deal with the question of whether there is a path that visits each edge exactly once and returns to the starting point (Eulerian circuit) or that passes each node exactly once and also returns to the starting node (Hamiltonian path problem).

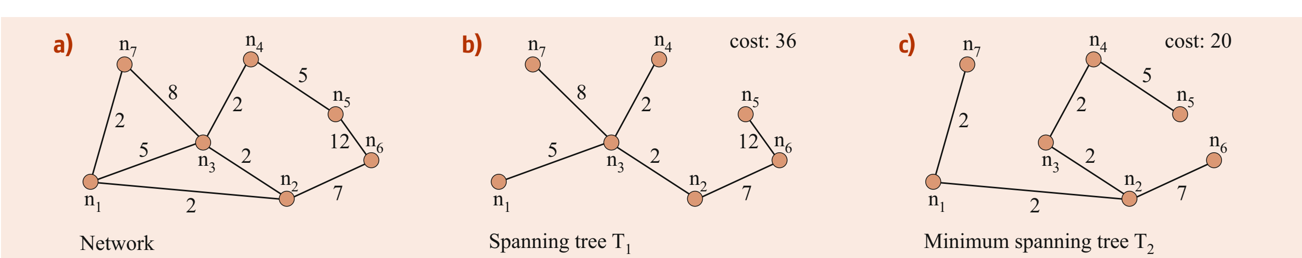


Fig. 3.42 Spanning trees

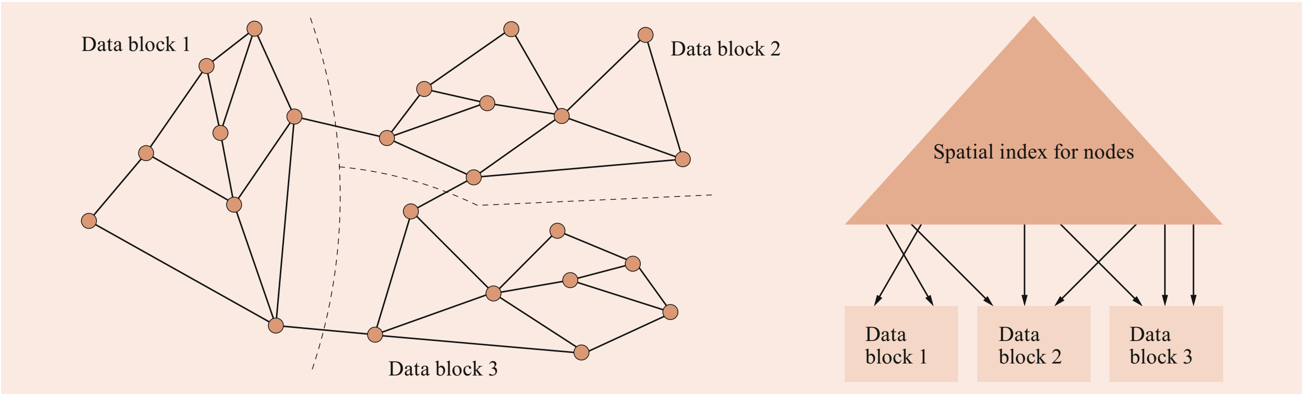
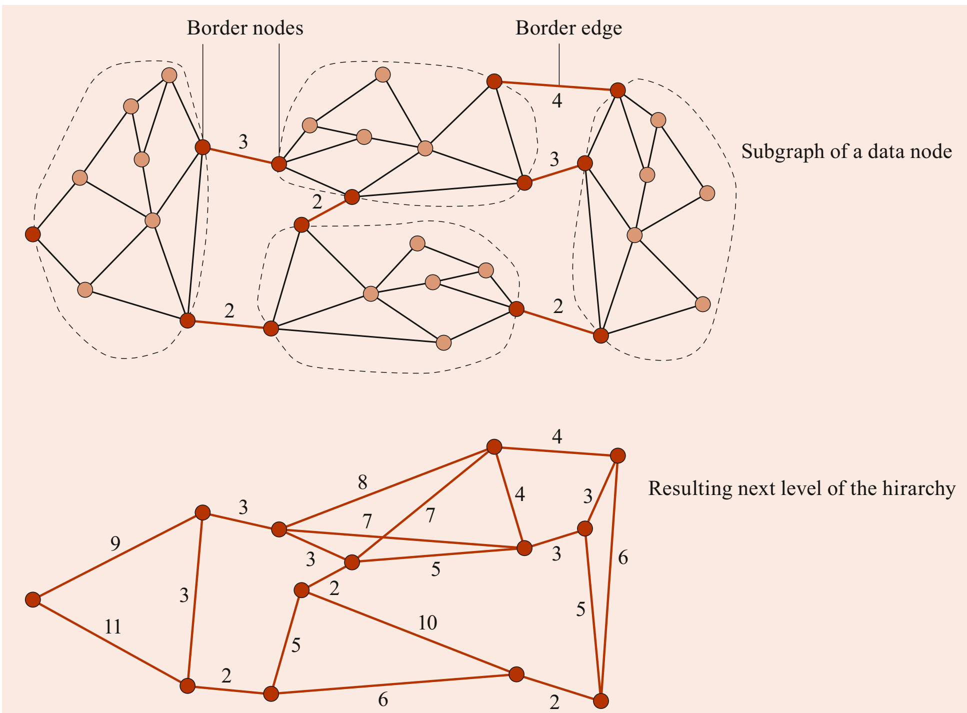


Fig. 3.44 Partitioning of networks

Fig. 3.45 Computing a network hierarchy



3.9.2 Partitioning and Hierarchies of Networks

To organize networks in spatial database systems, they must be – like alphanumerical and spatial data – partitioned to be stored in blocks in secondary storage. Because network analysis operations are performed in the main memory, the blocks required by the algorithm have to be read from the secondary storage.

Such a subdivision into subgraphs should consider spatial properties, as well as network aspects. An obvious spatial criterion is the proximity of nodes and edges. An essential network criterion is the reduction of the number of edges cut by the partitions. Figure 3.44 illustrates such an approach. Because the subdivision is not only based on the location of the nodes, the corresponding spatial index does not (completely) define the grouping of nodes and, therefore, works as a secondary index (indicated in the example by some close arrows pointing to different data blocks).

A partitioning of networks allows the building of *network hierarchies*. For the computation of the fastest connection between a starting place in Paris and a destination in Rome, e. g., it is not sensible to consider all minor streets in Lyon, Marseille, and Genoa. Instead, it is more reasonable to access the motorway in Paris and stay on it until Rome. Only around Paris and Rome do smaller roads need to be considered. This can be achieved by using network hierarchies.

After a partitioning of a network, all nodes of data block adjacent to nodes of another block are called *border nodes*; the edges between them are the *border edges*. Now, the shortest paths and distances between all border nodes can be computed for each subgraph. The border edges plus the edges representing the computed shortest paths form the next level of the network hierarchy. Figure 3.45 depicts a network subdivided into four subgraphs and the corresponding upper hierarchy level. The approach can be repeated recursively to compute further levels of the hierarchy.

The computation of the shortest path in a network hierarchy is done for the starting and the destination nodes in the detailed network until a border node is reached. Then, the computation can be performed using the smaller, upper-level network.

3.10 Raster Databases

The previous sections primarily discussed vector data. However, for many applications, raster data are also essential. One typical approach is to store the raster data outside the database system in files or to organize them using a geographical information system. In order to exploit the advantages of a database system (such as centralized data management, multiuser support, data security, transparent query processing, and the use of SQL) and provide open access, the use of database systems for raster data is becoming increasingly popular [7].

An obvious approach for storing raster data in a database system is the use of *binary large objects* (BLOBs). A BLOB can store and deliver arbitrary binary data. Because the database system does not know the syntax and semantics of the content, it cannot provide any specialized functionality or optimizations. If, in addition to the storage and simple access to raster data, more geospatial functions are needed, a spatial database system has to be extended using a raster data model and corresponding, raster-oriented methods. If primarily raster data is to be organized, the use of specialized raster database management systems is a suitable option.

3.10.1 Georeferencing and Spatial Selections

The *georeferencing* of a stored raster map is the most important requirement of spatial database systems dealing with raster data. Typically, the extent of a raster map is described by a rectangle or polygon, which is organized by a spatial index. For query processing, the index is used to determine the raster maps that belong to the result set.

If the database system stores the semantics of the raster data, content queries can also be processed. Such queries may refer to pixels (e.g., the raster belongs to the result set if x% of the raster cells or y neighboring pixels have a value below a given threshold) or use *world coordinates*

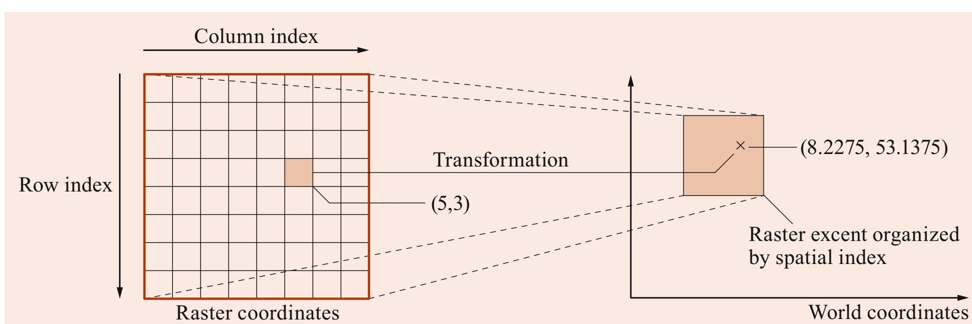


Fig. 3.46 Raster and world coordinates

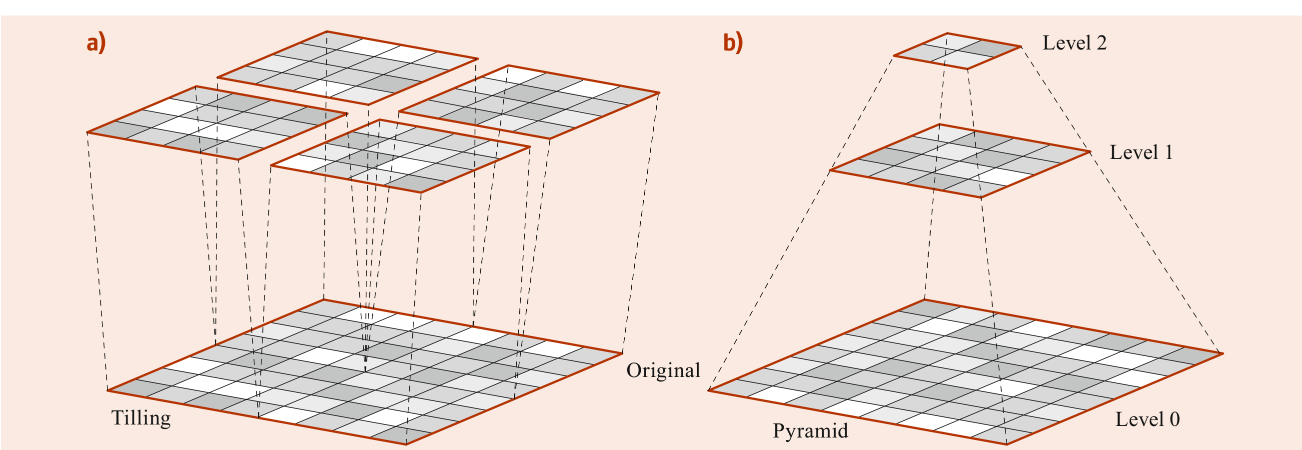


Fig. 3.47 Tiling (**a**) and pyramid (**b**)

(e. g., return the value of the raster cell at coordinate (8.2275, 53.1375)). In the second case, the spatial database system transforms coordinates between the world and the *raster coordinate system* (Fig. 3.46).

Depending on the types of raster maps, representations with or without color maps, matrices consisting of single or multiple bands and compressed or uncompressed raster storage have to be supported by the spatial database system.

3.10.2 Access Optimization

For large raster maps, *tiling* is appropriate. The raster matrix is subdivided into several subsets that can be retrieved and processed independently. Such tiling allows the transfer cost to be reduced when only a part of the raster map is requested. Figure 3.47a shows a tiling of a raster matrix into four smaller parts.

Applications typically require raster data at different resolutions. In the case of visualization, the requests depend on the current zoom factor. Therefore, a spatial database system should be able to provide raster maps at different resolutions. This can be achieved by a *pyramid* that stores a raster matrix at several resolutions. Typically, the number of pixels is halved per dimension between each pyramid level. To compute a raster matrix of smaller resolution, a resampling of the original raster data must be performed by the database system. Figure 3.47b illustrates a pyramid of three levels.

3.11 Spatiotemporal Databases

Current spatial database systems are able to organize the geometrical, topological, and thematic properties of geospatial features. In addition, *temporal* properties are another important category. For many spatial applications, the combined representation of space and time plays a crucial role. For instance, one would like

- To know when the Roman Empire had its largest extension
- To query all vehicles that drive faster than 120 km/h
- To determine the direction of hurricanes
- To inform all persons who will presumably pass a branch of the fast-food chain *X* in the next 15 min.

All four example queries show a combination of spatial and temporal query conditions. However, neither pure spatial nor pure temporal database systems offer suitable operations and data structures for such queries. This observation led to the development of *spatiotemporal database models*, fundamental data structures, and algorithms of *spatiotemporal databases* to organize and query data according to their spatial and temporal characteristics [35, Chap. 30].

3.11.1 Database Schema

An obvious solution is to organize time as an additional spatial-like dimension and, thus, to use it like a conventional spatial data type. However, such an approach does not consider the fact that time has other characteristics than spatial dimensions and that the combination of space and time raises special questions. Discrete modification, such as the subdivision of a cadastral parcel, should be distinguished from continuous changes. While discrete changes can still be handled relatively well with conventional database techniques, it is more difficult to represent continuous changes. Continuous changes are often involved with movements. Many geospatial objects – humans, vehicles, but also geographical objects – move continuously, i. e., they change their position and/or their shape. If only the changes of an object's position are of interest, the object is considered as a *moving point*, even if it is not a point. For example, these positions can belong to vehicles of a fleet or to animals on their trails. In other cases, the change of the shape of geospatial features, such as melting glaciers and draining lakes, is of interest. Then, the abstraction to a single point is not appropriate.

Data Types for Time

ISO 19108 *Geographic information – Temporal schema* deals with the representation of temporal data. This standard defines two *temporal primitives* (*TM_Primitive*) for the description of time: the *instant* (*TM_Instant*), representing a point (*position*) in time, and the *period* (*TM_Period*), representing an extent in time. The period is limited by two instances (beginning and ending). For open periods, the beginning or the ending is indeterminate. A special case of *TM_Instant* is *TM_Coordinate*; in this case, the instant is described by a single numerical value that represents the distance from an origin.

ISO 19108 specifies operations for computation of the relative order between two temporal primitives (e. g., before, after, contains). The data type *TM_Duration* represents the *duration* of a period or a temporal distance between two instances (e. g., *P5DT4H30.7M* corresponds to 5 days, 4 h, 30.7 min). In addition, the standard defines calendars as *temporal reference systems*.

Spatiotemporal Data Types

ISO 19141 *Geographic information – Schema for moving objects* provides a data model for *moving objects*, i. e., for features whose locations change over time. It is limited to features without deformations but allows rotations. Such *temporal geometries* are extensions to feature geometries as defined by *ISO 19107 Spatial Schema*.

Trajectories are curves that represent the path of a point in the temporal geometry. They are controlled by a parameter value from an interval; in the case of a *temporal trajectory*, this value is an instant. Figure 3.48a illustrates temporal tra-

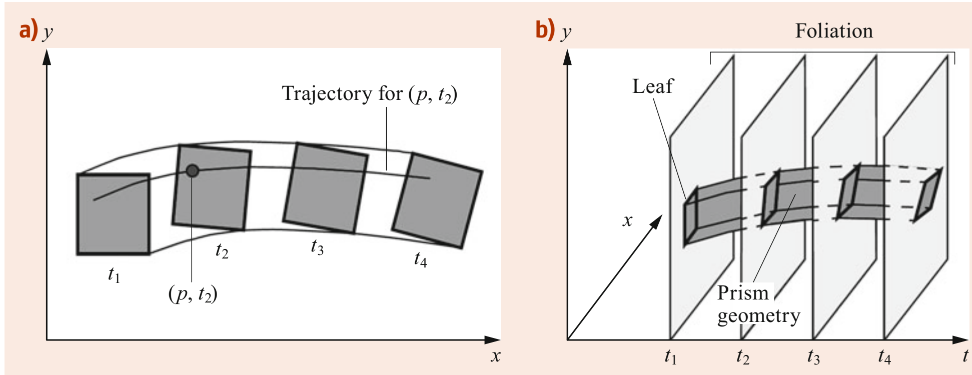


Fig. 3.48 Temporal trajectories of a moving rectangle **(a)** and the foliation and prism geometry of this moving rectangle **(b)**

jectories of a moving rectangle. The temporal geometry at a single instant is called *leaf*. A set of instants leads to a *foliation*; the resulting leaves span a *prism geometry*. Using the same set of instants, each trajectory of a moving object is included in the prism geometry. Figure 3.48b shows the foliation and prism geometry of a moving rectangle.

For a temporal geometry, ISO 19141 specifies several operations. Some examples are the following:

- The functions *startTime* and *endTime* that return the start and end time of a movement as a *TM_Coordinate*.
- The function *leafGeometry* that provides a leaf as ISO 19107 geometry for a given *TM_Coordinate*.
- The function *trajectory* that computes the temporal trajectory for a given position and *TM_Coordinate*.

For a temporal trajectory, the start and end time of its movement, its geometry as a curve, the position for a given instant, the set of temporal primitives for a given position, as well as the velocity and acceleration at a given *TM_Coordinate* can be computed.

Another popular database model for the representation of moving objects was proposed by *Güting* et al. [36].

3.11.2 Basic Queries

A spatiotemporal database model enables the modeling of spatiotemporal data and formulation of queries that retrieve the desired data from a database. For efficient processing of such queries, a spatiotemporal database system provides a set of *spatiotemporal basic queries*.

The window query is one of the most important spatial basic queries. Three variants exist if this concept is generalized for spatiotemporal databases [37].

- A *timeslice query* determines, for a query window w and a time instant t, all moving objects whose geometry intersects w at time t.

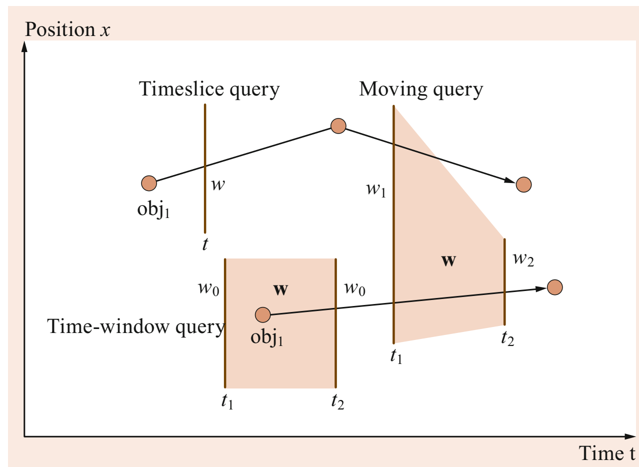


Fig. 3.49 Three spatiotemporal variants of a window query for the one-dimensional case

- A *time-window query* computes, for a static query window w and a period p, all moving objects whose geometry intersects the window $at(w,t)$ at instant $t \in p$. Since w does not change, it holds that $at(w,t) = w_0 \in$ Envelope.
- A *moving query* determines, for a moving query window w and a period p, all moving objects whose geometry intersects w at instant $t \in p$. The change of the query window can be linearly interpolated between two windows w_1 and w_2 ($w_1, w_2 \in$ Envelope).

Figure 3.49 depicts examples of these three variants of a window query for the one-dimensional case, i. e., the window is a one-dimensional (1-D) interval.

A *spatiotemporal join* has a spatiotemporal join condition. An example would be the retrieval of all pairs of moving objects that intersect during a given period. The *spatiotemporal nearest-neighbor query* computes the k nearest moving objects. In the simplest case, the query object is a point at a given point of time. In more complex cases, the query object is a moving object during a given period.

3.11.3 Spatiotemporal Indexing

Like for other types of databases, indexing is an essential precondition for efficient processing of spatiotemporal queries. A *spatiotemporal index structure* may target one of the following objectives.

- Indexing of the past: the query time is before the current time, so that all spatial and temporal information is accessible from the database.
- Indexing of the present: all queries use the current time.
- Indexing of the (near) future: such *predictive queries* are processed using the known former position, speed, and direction of moving objects that are stored in the database and extrapolating them into the future.

Depending on the objective, different spatiotemporal index structures have been designed.

TPR-Tree

The time-parameterized R-tree (*TPR-tree*) is a variant of the R-tree for organizing moving objects [37]. The TPR-tree supports spatiotemporal queries concerning the current or future position of the stored objects.

The TPR-tree stores the position $\mathrm{pos}(t_\mathrm{r})$ and the vector $\boldsymbol{v}(t_\mathrm{r})$ representing the direction and velocity of an object-specific point in time t_r, for every moving object. A two-dimensional point p with $\mathrm{pos}(t_\mathrm{r})$ and $\boldsymbol{v}(t_\mathrm{r})$ can be described by two coordinates (x, y) and a pair $(\Delta x, \Delta y)$ indicating the amount of change in x and y during a unit of time (Fig. 3.50). For any instant $t \geq t_\mathrm{r}$, position and velocity are calculated as follows (if the point p is not modified in the meantime by a database operation):

$$\mathrm{p.pos}(t) = (x + (t - t_\mathrm{r}) \cdot \Delta x, y + (t - t_\mathrm{r}) \cdot \Delta y),$$
$$\mathrm{p}.v(t) = \mathrm{p}.v(t_\mathrm{r}) = (\Delta x, \Delta y).$$

As a variant of the R-tree, the TPR-tree must be able to compute a minimum bounding rectangle around moving points. A *time-parameterized minimum bounding rectangle* (TP-MBR) consists of a set of moving points p_i for a lower and an upper bound for the position and for the movement. For a point in time t, the attributes of a TP-MBR can be computed for the x-dimension as

$$x_{\min}(t) = \min\{\mathrm{p}_i.\mathrm{pos}(t).x\},$$
$$x_{\max}(t) = \max\{\mathrm{p}_i.\mathrm{pos}(t).x\},$$
$$\Delta x_{\min}(t) = \min\{\mathrm{p}_i.v.\Delta x\},$$
$$\Delta x_{\max}(t) = \max\{\mathrm{p}_i.v.\Delta x\}.$$

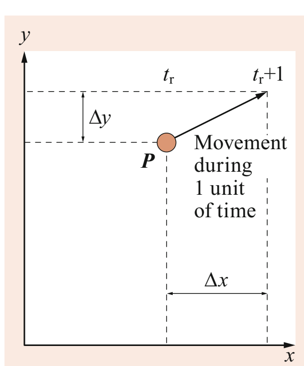


Fig. 3.50 Representation of moving points

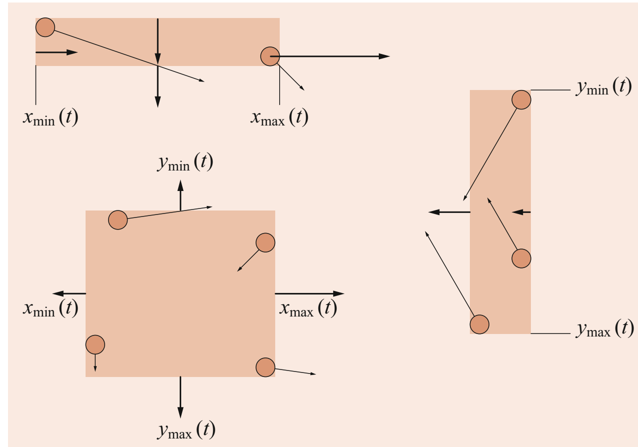


Fig. 3.51 Computation of TP-MBRs

The y-dimension for the attributes is calculated correspondingly. Figure 3.51 illustrates the computation of TP-MBRs. The TP-MBR for a set of TP-MBRs can be computed similarly.

A TPR-tree stores TP-MBRs instead of normal rectangles for describing block regions. In principle, the TPR-tree uses the same algorithms as R-trees or R*-trees. The *choose-subtree* and *split* operations (Sect. 3.8.4) require geometric algorithms for rectangles such as the computation of area, perimeter, distance to another rectangle, and intersection with another rectangle. Computing these properties for TP-MBRs requires consideration of an actual point in time or an actual period. In the first case, a TP-MBR is mapped to a rectangle. In the second case, a polygon – the so-called *sweeping region* (Fig. 3.52a) – is the result.

For efficient support of predictive queries, a TPR-tree introduces an application-dependent *horizon*, h, which indicates how far the tree should *see* into the future. The choose-subtree and split operations use the period $[t_\mathrm{curr}, t_\mathrm{curr} + h]$ to compute their geometric algorithms (t_curr is the current time). Thus, they are performed on sweeping regions. Experimental investigations have shown that h should be within the interval $[0.5 \cdot \Delta\mathrm{upd} + \Delta\mathrm{qry}, \Delta\mathrm{upd} + \Delta\mathrm{qry}]$, where $\Delta\mathrm{upd}$ denotes the average period between two updates of a moving object and $\Delta\mathrm{qry}$ is the average period that queries look into the future.

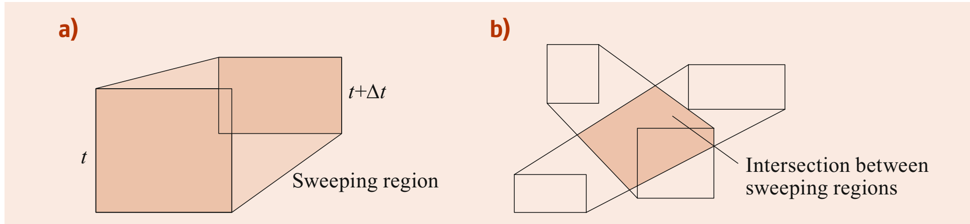


Fig. 3.52 Sweeping regions

The spatiotemporal basic queries presented in Sect. 3.11.2 can be performed by a TPR-tree. In the case of time-window queries, the query window must be intersected with sweeping regions computed from the TP-MBRs. For moving queries, the query window, as well as the TP-MBRs, are sweeping regions for which an intersection must be determined (Fig. 3.52b).

3.12 Spatial Database Systems

In the past decade, the support of spatial data became a common feature of database management systems. Relational client-server systems like *IBM DB2*, *IBM Informix*, *Microsoft SQL Server*, *MySQL*, *Oracle Database*, and *PostgreSQL/-PostGIS* provide a spatial data model, analysis functions, and spatial query and index capabilities [7]. All these systems support (more or less) the simple feature model of OGC/ISO 19125. EPSG codes are used to identify spatial reference systems. Oracle and PostgreSQL/PostGIS also facilitate parts of SQL/MM Spatial. Both systems provide extensions for linear referencing systems, network, and 2-D topological class models, and georeferenced raster data. Often, R-trees are implemented to accelerate spatial queries, but also hierarchical grid indexes are used (Microsoft SQL Server).

Not only traditional relational client-server systems work as spatial databases. *SAP HANA* is an in-memory, column-oriented, relational database server that can handle spatial data according to the simple feature model. This statement also holds for embedded databases that are tightly integrated with applications and provide on-disk and in-memory storage (and combinations of both). *SpatiaLite* is an extension of the embedded database *SQLite* that supports the simple feature model and spatial query processing. The *H2* database engine achieves the same by incorporating the Java programming library *JTS*.

NoSQL databases have gained increased importance in recent years. Some of them also provide spatial functionality. For example, the document-oriented system *MongoDB* allows storing GeoJSON data and supports a spatial index. However, its analysis and query capabilities are (currently) rather limited compared to relational spatial database systems. The graph database *Neo4j* also includes common geometry types and an R-tree. In addition to the support of basic spatial queries, spatial operators can be applied while traversing the graph.

References

1. Zölitz, R., Hartleib, J.: GeoGREIF Geographische Sammlung (Ernst-Moritz-Arndt-Univ., Greifswald). http://greif.uni-greifswald.de/geogreif/, Accessed 23 May 2010
2. Worboys, M., Duckham, M.: GIS: a computing perspective, 2nd edn. CRC Press, Boca Raton (2004)
3. Steel, T.B.: Database standardization. A status report. In: Proc. 1975 ACM SIGMOD, pp. 149–156. (1975)
4. Codd, E.F.: A relational model of data for large shared data banks. Commun. ACM **13**(6), 377–387 (1970)
5. Pin-Shan Chen, P.: The entity-relationship model-toward a unified view of data. ACM Trans. Database Syst. **1**(1), 9–36 (1976)
6. Chen, P.P.: Entity-Relationship Modeling: Historical Events, Future Trends, and Lessons Learned. Louisiana State Univ., Baton Rouge (2010). https://bit.csc.lsu.edu/~chen/pdf/Chen_Pioneers.pdf (Accessed 23 May 2010)
7. Brinkhoff, T.: Geodatenbanksysteme in Theorie und Praxis, 4th edn. Wichmann, Heidelberg (2022)
8. Litwin, P.: Fundamentals of relational database design, adapted from the design chapter of *Microsoft Access 2 Developer's Handbook*. http://r937.com/relational.html, Accessed 23 Feb 2010
9. Robbins, R.J.: Database Fundamentals. Johns Hopkins Univ., Baltimore (1995). http://www.esp.org/db-fund.pdf (Accessed 23 May 2010)
10. Pascal, F.: SQL and Relational Basics. M&T Books, New York (1990)
11. GeekInterview: A venture from Exforsys.com. What is normalization?. http://www.geekinterview.com/question_details/15914, Accessed 08.2019
12. MongoDB: https://www.mongodb.com/, Accessed 30 December 2018
13. JCC Consulting: SQL Standards Page. JC Consulting, Granville (2010). available from http://www.jcc.com/sql.htm (Accessed 2019)
14. FFE Software: SQL Tutorial, pure SQL92. FFE Software Inc., El Cerrito (2005). available from http://www.firstsql.com/tutor.htm (Accessed 2019)
15. Riordan, R.M.: Designing Relational Database Systems. Microsoft Press, Redmont (1999). http://technet.microsoft.com/en-us/library/cc966426.aspx (Accessed 2019)
16. Heron, T.: CS319: Theory of Databases OODBMS, ORDBMS. Univ. of Warwick, Coventry (2004). http://www.dcs.warwick.ac.uk/~theron/cs319/oo-slides.pdf (Accessed 9 May 2010)
17. Plümer, L.: Geoinformation – Neue Medien für die Einführung eines neuen Querschnittsfaches (2002). http://www.geoinformation.net
18. Atkinson, M., Bancilhon, F., DeWitt, A.D., Dittrich, K., Maier, D., Zdonik, S.: The Object-Oriented Database System Manifesto (Computer Based Learning Unit, Univ. of Leeds) (1995).

http://www.cs.cmu.edu/afs/cs.cmu.edu/user/clamen/OODBMS/Manifesto/htManifesto/Manifesto.html, Accessed 2019
19. Maier, D.: Object-oriented database theory: an introduction. Seminar held at the Technical University of Munich. http://www.csd.uoc.gr/hy562/Papers/OODBMS.pdf, Accessed 30 July 2011
20. Developer Shed: Introduction to RDBMS, OODBMS and ORDBMS (2010). http://www.aspfree.com/c/a/Database/Introduction-to-RDBMS-OODBMS-and-ORDBMS/3/, Accessed 9 Mar 2010
21. Google Cloud: https://cloud.google.com/bigtable/, Accessed 30 December 2018
22. Rigaux, P., Scholl, M., Voisard, A.: Spatial Databases, With Applications to GIS. Morgan Kaufmann, San Francisco (2002)
23. Shekhar, S., Chawla, S.: Spatial Databases – A Tour. Prentice Hall, Upper Saddle River (2003)
24. Brinkhoff, T., Kriegel, H.-P., Schneider, R., Seeger, B.: Multi-step processing of spatial joins. In: Proc. ACM SIGMOD, Int. Conf. Manag. Data, Minneapolis, pp. 209–220. ACM, New York (1994)
25. Papadopoulos, A.N., Manolopoulos, Y.: Nearest Neighbor Search – A Database Perspective. Springer, Berlin, Heidelberg, New York (2005)
26. Orenstein, J.A.: Redundancy in spatial databases. In: Proc. ACM SIGMOD Int. Conf. Manag. Data, Portland, pp. 294–305. ACM, New York (1989)
27. Brinkhoff, T., Kriegel, H.-P., Schneider, R.: Comparison of approximations of complex objects used for approximation-based query processing in spatial database systems. In: Proc. 9th Int. Conf. Data Eng., Vienna, pp. 40–49. IEEE Computer Society, New York (1993)
28. Seeger, B., Kriegel, H.-P.: Techniques for design and implementation of efficient spatial access methods. In: Proc. 14th Int. Conf. Very Large Data Bases, Los Angeles, pp. 360–371. VLDB Endowment, Almaden (1988)
29. Samet, H.: Foundations of Multidimensional and Metric Data Structures. Morgan Kaufmann, San Francisco (2006)
30. Nievergelt, J., Hinterberger, H., Sevcik, K.C.: The grid file: an adaptable, symmetric multikey file structure. ACM Trans. Database Syst. **9**, 38–71 (1984)
31. Guttman, A.: R-trees: a dynamic index structure for spatial searching. In: Proc. ACM SIGMOD Int. Conf. Manag. Data, Boston, pp. 47–57. ACM, New York (1984)
32. Manolopoulos, Y., Nanopoulos, A., Papadopoulos, A.N., Theodoridis, Y.: R-Trees – Theory and Applications. Springer, London (2006)
33. Beckmann, N., Kriegel, H.-P., Schneider, R., Seeger, B.: The R*-tree: An efficient and robust access method for points and rectangles. In: Proc. ACM SIGMOD Int. Conf. Manag. Data, Atlantic City, pp. 322–331. ACM, New York (1990)
34. Kamel, I., Faloutsos, C.: Hilbert R-tree: An improved R-tree using fractals. In: Proc. 20th Int. Conf. Very Large Data Bases, Santiago de Chile, pp. 500–509. VLDB Endowment, Almaden (1994)
35. Güting, R.H., Schneider, M.: Moving Objects Databases. Morgan Kaufmann, San Francisco (2005)
36. Güting, R.H., Böhlen, M.H., Erwig, M., Jensen, C.S., Lorentzos, N.A., Schneider, M., Vazirgiannis, M.: A foundation for representing and querying moving objects. ACM Trans. Database Syst. **25**, 1–42 (2000)
37. Šaltenis, S., Jensen, C., Leutenegger, S.T., Lopez, M.A.: Indexing the positions of continuously moving objects. In: Proc. ACM SIGMOD Int. Conf. Manag. Data, Dallas, pp. 331–342. ACM, New York (2000)

Thomas Brinkhoff Thomas Brinkhoff is Professor of Geoinformatics at Jade University Oldenburg, Germany. He studied computer science at the University of Bremen and received his doctoral degree from the University of Munich. His research interests are spatial and spatiotemporal database systems, mobile GIS, location-based services, and the Geospatial Sensor Web. He published the first German textbook on spatial database systems.

Wolfgang Kresse Wolfgang Kresse holds a Diploma degree in Geodesy and a Doctor's degree in Digital Cartography, both from the University of Bonn, Germany. He was Professor of GIS and Photogrammetry at the University of Applied Sciences in Neubrandenburg, Germany, and Professor of remote sensing at the Politechnika Koszalińska in Poland. As a Visiting Professor he was affiliated with the Canada Centre for Remote Sensing in Ottawa (2002/2003) and the University of Maine, USA (2008/2009). He has also worked as a software developer and customer trainer at the Carl Zeiss company with a focus on graphics, photogrammetry, and data exchange.

3

Encoding of Geographic Information

4

Clemens Portele

Contents

Abstract

Encoding is the process of coding geographic information into a system-independent data structure suitable for transport and storage. The inverse process is called *decoding*. The resulting data structure may be stored on digital media or transferred over a network using transfer protocols. While the data structure is intended to be read and interpreted by computers, it may be in a form that is human-readable.

For cases where the data is consistent with an application schema or a taxonomy (Chap. 1), rules are needed for an unambiguous encoding and decoding of data without loss of information. Such encoding rules decouple the representation of data during transport or storage from the representation for other purposes, such as spatial analysis, presentation to the user, etc.

Wikipedia describes *encoding* as *the process that converts information from a source into symbols for communication or storage*. The reverse process is *decoding*, which converts the sequence of symbols *back into a form that the recipient understands*.

A difference between the two purposes, communication and storage, is that encoding for storage is typically a process that is internal to a single system, while encoding for communication is a process that involves at least two parties. An example of storing encoded data is an application that uses a database to store and query information about buildings and cadastral parcels. An example of communicating encoded data is the exchange of building information as a file; one party exports the building information in a file and sends it to a second party that reads and processes (decodes) the file.

This chapter is mainly about encoding for communication.

Geographic data is typically encoded in units of features, which are the real-world entities of interest, as discussed in Chap. 1. The spatial data on the Web Best Practices of W3C and OGC [1] use the term "spatial things" for features.

There are different approaches to encoding feature-based data, often linked to the different approaches how geographic information is modeled, as discussed in Chap. 1.

On one end of the scale, there is the strict schematic approach, where each feature is of a certain feature type, and data can be fully validated against an application schema. This is the approach that has been promoted in most of the standards by OGC or ISO/TC 211. GML [2, 3], GeoPackage [4], and Shapefiles [5] are examples of this approach to encoding geographic data.

The strict schematic approach is best for cases where the provider and the receiver of data benefit from a strong contract, i. e., a deep common understanding of the real world. It typically requires software developed for the relevant application schema(s) – or software that is able to parse application schemas in a generic way.

At the other end of the scale are schemaless, dynamic encodings where any feature can take any property and may be associated with none, one, or multiple feature types. Geo-

C. Portele (✉)
interactive instruments GmbH
53115 Bonn, Germany
e-mail: portele@interactive-instruments.de

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_4

JSON [6] is an example of an encoding that supports this approach.

One approach that draws from both of these extremes is encodings that support any property and multiple feature types, but where the feature types and properties are formalized and registered in a vocabulary. The feature model of GeoSPARQL [7] is a typical example of this approach.

All of these encodings, however, convert the features with their properties in a machine-processable way. The difference is in how rigidly they specify the properties that a feature has or may have. This chapter will mainly focus on schema-based encodings, which can be defined using so-called encoding rules.

The use of well-defined encoding rules supports the interchange of data in multiple representations; for example, a data provider may offer its data encoded in formats such as GeoJSON [6], the Geography Markup Language (GML) [2, 3], or Shapefiles [5] (commonly used formats are explained in more detail below). The encoding rules enable a receiver to decode the data and understand it in the context of the application schema that defines the structure of the data on a conceptual level – independently of the encoding actually used in the transfer process.

This approach is consistent with the model-driven architecture (MDA) approach to information technology (IT) systems' development fostered by the Object Management Group (OMG), which is based on forming a separation between the specification of the essential functionality as a platform-independent model (PIM) – in this case, the application schema – and the realization using a more detailed and platform-specific model (PSM) – in this case, the data structure schema describing the encoding. See Chap. 1 of this Handbook for more details.

The relevant concepts for the interchange of geographic data between systems are specified in ISO 19118 [8]. Some encoding rules have been standardized; for example, by GML (ISO 19136 *Geography markup language*) [2, 3] and ISO/TS 19139-1 *Metadata – XML schema implementation* [9]. An example is used to illustrate how data can be encoded in different encodings that are commonly used to encode geographic data, including GML, GeoJSON, and Shapefiles.

4.1 Encoding Concepts

The encoding process is discussed in detail in ISO 19118 *Encoding* [8]. This section provides a description of the fundamental concepts of data interchange from this standard. The content found in this section is partially derived from clause 6 of ISO 19118 but has been amended to describe the process in cases where no application schema is specified.

4.1.1 Data Interchange

An overview of a data interchange is shown in Fig. 4.1. System A wants to send geographic data to system B. Both systems, A and B, store data in an internal database according to an internal schema, but the schemas are usually different. The following logical steps are taken in order to transfer geographic data from A's internal database to B's internal database.

1. System A first translates its internal data into a data structure that is according to an application schema I. This is done by defining a transformation from the concepts of the internal schema to the concepts of the application schema and by applying appropriate software to transform the data. In Fig. 4.1 this mapping is denoted M_{AI}. The result is an application-schema-specific but system-dependent data structure i_{A}.
2. System A then uses an encoding service, which applies the encoding rule R to create a data structure d that is system independent and, therefore, suitable for transfer. It may be stored in a file system or transferred using a transfer service.
3. Using a transfer service, the data structure d is sent to system B. The transfer service follows a transfer protocol and handles packaging, compression, encryption, how the actual transportation over an online or offline communication medium takes place, etc.
4. The transfer service on system B receives the transferred data, and according to the protocol used it is unpacked and stored as the system-independent data structure d, e.g., as an intermediate file.
5. In order to get an application-schema-specific data structure i_{B}, system B applies the inverse encoding rule R^{-1} to decode the encoded data.
6. To use the dataset, system B translates the application-schema-specific data structure i_{B} into its internal database. This is done by defining a transformation from the concepts of the application schema to the concepts of the internal schema and by applying appropriate software to transform the data. In Fig. 4.1, this mapping is denoted M_{IB}.

So, to ensure a successful interchange, A and B must agree on three aspects:

- An application schema I that defines the semantics of the content and the logical structures of the geographic information. See Chap. 1 for more details on application schemas.
- An encoding rule R defining the conversion rules for how to code geographic information corresponding to an application schema into a system-independent data structure. This chapter discusses such encoding rules in more detail.

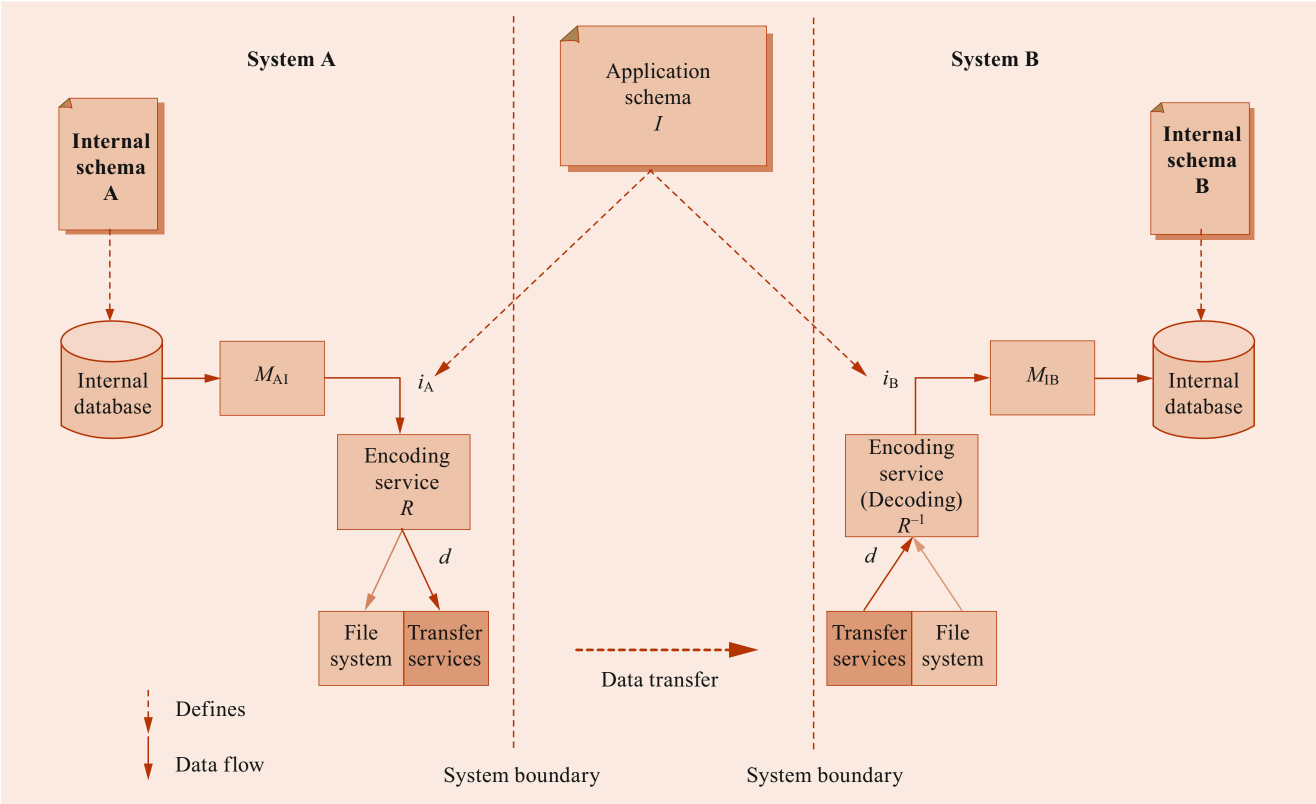


Fig. 4.1 Overview of data interchange between two systems [8]

- A transfer protocol for the transfer of the data between the two systems. A typical example for such a transfer protocol is the hypertext transfer protocol (HTTP) – or Web APIs that are based on HTTP, like the emerging OGC API family of standards.

On the web, encoding will typically be negotiated between a web client and the web server as part of the HTTP protocol based on the encodings that a client can understand and the encodings that a server can provide (*content negotiation*). In this process, the encodings are identified using internet media types; for example, the media type *application/geo+json* identifies a GeoJSON document [10], *application/gml+xml* a GML document [11], and *application/geopackage+sqlite3* a GeoPackage database file [12].

Note that the description above assumes that both A and B use schema-based approaches for their internal storage of features. Depending on the extent to which the systems internally use a dynamic feature storage, the need for an application schema may be less important or not important at all. In that case, data from system A may be directly ingested into system B. Then, the data interchange process no longer has a responsibility to *understand* the data from system A. Instead, this is now the responsibility of every process that uses the data in system B.

4.1.2 Encoding Rules

An encoding rule is an identifiable collection of conversion rules that defines the encoding for a particular data structure. It specifies the data types to be converted, as well as the syntax, structure, and coding schemes used in the resulting data structure. An encoding rule is applied to application-schema-specific data structures to produce system-independent data structures suitable for transport or storage. In order to define an encoding rule, three important aspects are specified

- The input data structure
- The output data structure
- The conversion rules between the elements of the input and the output data structures.

Concept

The schemas of both the input and output data structures are written using a schema language, and the concepts in the language are used to define the encoding rule.

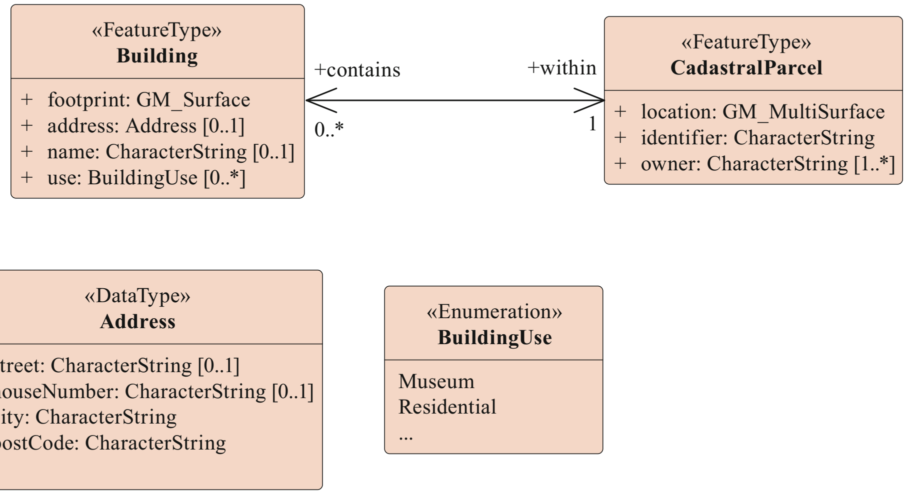


Fig. 4.2 UML class diagram

Role of the light-brown insets

To illustrate the encoding concepts, examples using GeoJSON, GML, and SQL are included in this chapter. These parts are shown in boxes with light-brown background.

In the examples, we will use a simple application schema with feature types for buildings and cadastral parcels. The UML class diagram in Fig. 4.2 describes the two feature types and their properties.

Input Data Structure

The input data structure is an application-schema-specific data structure. The data structure can be thought of as a set of instances, i. e., $i = \{i_1, \ldots, i_p\}$, where each instance i_i is an instance of a concept I_j defined in the application schema I. The application schema defines a set of concepts defined in the application schema $I = \{I_1, \ldots, I_m\}$.

The application schema is described in a formal way, often using a conceptual schema language, for example, unified modeling language (UML). The ISO standards ISO 19109 *Rules for application schema* and ISO 19103 *Conceptual schema language* specify a UML profile for describing application schemas for datasets of geographic information.

In practice, other UML profiles, other conceptual schema languages, taxonomies and controlled vocabularies, or less formal ways of documenting a schema are in frequent use too, but these are only partly considered in this chapter.

Output Data Structure

The output data structure is defined by a data structure schema $D = \{D_1, \ldots, D_s\}$; D is the schema for the output structure and is not shown in Fig. 4.1. The output data structure can be thought of as a set of instances, i. e., $d = \{d_1, \ldots, d_q\}$, where each instance d_i is an instance of a concept D_j. The schema D defines the syntax, structure, and coding schemes of the output data structure.

While the data structure is intended to be read and interpreted by systems as described above, it may be in a form that is human-readable.

The data structure schema is described using a schema language suitable for such data structures, for example, XML Schema [13], JSON schema [14] or SQL DDL [15].

Schemas for feature type "Building" for GML and GeoJSON

The following XML schema snippet specifies the schema components for a building feature in GML. Other schema components from GML (for example, the Surface geometry) and from the GML application schema (for example, the BuildingUse enumeration) are referenced. Building instances can be validated against this schema definition.

```
<xs:element name="Building"
            type="app:BuildingType"
            substitutionGroup=
             "gml:AbstractFeature"/>
<xs:complexType name="BuildingType">
 <xs:complexContent>
  <xs:extension base="gml:AbstractFeatureType">
   <xs:sequence>
    <xs:element name="within"
                type=
                 "app:CadastralParcelPropertyType"/>
    <xs:element name="address"
                minOccurs="0"
                maxOccurs="1"
                type="app:AddressPropertyType"/>
    <xs:element name="footprint"
                type="gml:SurfacePropertyType"/>
    <xs:element name="name"
                type="xs:string"
                minOccurs="0"
                maxOccurs="1"/>
    <xs:element name="use"
                minOccurs="0"
                maxOccurs="unbounded"
                type="app:BuildingUseType"/>
   </xs:sequence>
```

```
    </xs:extension>
  </xs:complexContent>
</xs:complexType>
```

Today, it is not common practice to specify feature type definitions in JSON schema. GeoJSON as the most commonly used JSON encoding for feature data specifies a general JSON structure for feature data (for which a JSON Schema exists, see http://geojson.org/schema/Feature.json), but it is currently not common practice to specify a schema for the feature properties. The properties are treated as key-value pairs and are not validated against a schema. A key-value pair is an item of data that is identified by a unique name. The key is the name, and the value is the content.

Conversion Rules

A conversion rule specifies how an instance in the input data structure is converted to zero, one, or more instances in the output data structure. The conversion rules are based on the concepts of the conceptual schema language, on the one hand, and on the concepts of the data structure schema language, on the other hand.

To cover all potential cases, an encoding rule has to specify a conversion rule R_i for each of the legal combinations of concepts in the conceptual schema language. The set of conversion rules is $R = \{R_1, \ldots, R_n\}$, where R_i is the i-th conversion rule.

As a result, conversion rules may be defined based on the two schema languages and not on any particular application schema. This is a generic approach that allows developers to write application-schema-independent encoding services, which can be used for different application schemas, as long as the schemas are defined in the same conceptual schema language.

In practice, however, encoding rules are often specified less generically or only implicitly.

- Implicitly. Special, but common, cases are output data structure schemas with no explicitly specified application schema, i. e., the application schema can only be implicitly derived from the output data structure schema by straightforward rules. A result of this is that such application schemas are limited to the expressiveness of the respective data structure schema language. A common example is *data formats* that are specified directly as GeoPackages or Shapefiles. From these files, an application schema may be derived implicitly.
  In *data formats* that are schemaless key-value pairs as supported, for example, by JSON, there may not be any predefined schema for the encoded data structures. In those cases, the receiving system may derive a schema from the keys and values found in the data ("schema on read").
- Less generically. Another common case is that output data structure schemas are specified based on the concepts in the application schema, not the concepts of the conceptual schema language. A typical reason for such an approach is optimizations in the encoded data structures.
  The GML schema, i. e., the XML schema components specified in the GML standard, is an example. While some parts of the GML schema are derived using the GML encoding rule for application schemas (Sect. 4.2.1), other parts of the GML schema have been created manually from the underlying conceptual schemas of the ISO 19100 series. For example, to enable a compact encoding of coordinates, the PointArray type from the spatial schema in ISO 19107 has been implemented in an optimized XML encoding. Details of these cases are documented in annex D of GML [2].

4

Encoding a curve geometry in GeoJSON, GML, and SQL

Geometries have optimized encodings on most feature representations. Here is a simple example of a curve with linear interpolation connecting two points. The encoding is compact in all cases, with the coordinates taking up most of the characters.

In GeoJSON:

```
{
   "type": "LineString",
   "coordinates": [ [ 6.3962505824686255,
                      50.352934866558016 ],
                    [ 6.396334132914268,
                      50.353050797770116 ] ]
}
```

In GML:

```
<gml:LineString>
 <gml:posList> 6.3962505824686255
              50.352934866558016
               6.396334132914268
              50.353050797770116
 </gml:posList>
</gml:LineString>
```

In SQL using Well-Known Text (WKT) [33, 34]:

```
LINESTRING( 6.3962505824686255
           50.352934866558016,
            6.396334132914268
           50.353050797770116)
```

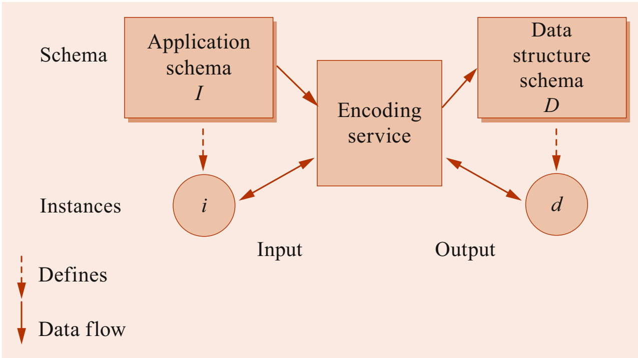


Fig. 4.3 Overview of the encoding process [8]

Encoding Service

An encoding service is a software component that implements an encoding rule and provides an interface to an encoding and/or decoding functionality. It is an integral part of data interchange.

Figure 4.3 presents the details of an encoding service and its relationships to important specification schemas (application schema I and data structure schema D). The encoding service reads the input data structure i and converts the instances to an output data structure d, or vice versa.

The encoding service converts from a structure that is already application schema specific to a representation suitable for transfer, i. e., the internal schemas A and B will usually specify more information than is represented in the application schema I. For example, a system A will often store additional information used for the maintenance of the data but that is not relevant for users. As a result, the internal databases in systems A and B will usually contain more information than will be transferred based on the application schema.

Example of on output data structure in GeoJSON and GML

The following examples illustrate how the same building instance (item i in Fig. 4.3) can be expressed in GeoJSON, GML, and SQL (item d in Fig. 4.3). We use the birthplace of Ludwig van Beethoven in the example.

The building in GeoJSON:

```
{
  "type": "Feature",
  "id": "DENW36AL10006wORBL",
  "geometry": {
    "type": "Polygon",
    "coordinates": [ [
              [ 6.77145389740199,
               51.513055921535326 ],
              [ 6.771456210102157,
               51.51303866250359 ],
              [ 6.77167501551265,
               51.51303808155525 ],
              [ 6.771674222426856,
               51.51303897479996 ],
              [ 6.771675139403735,
               51.513089282405375 ],
              [ 6.771453749979931,
               51.51308846525745 ],
              [ 6.77145389740199,
               51.513055921535326 ]
      ] ]
    },
  "properties": {
    "name": "Beethovenhaus",
    "type": "Building",
    "use": [ "Museum" ],
    "address": {
      "streetName": "Bonngasse",
      "houseNumber": "20",
      "city": "Bonn",
      "postCode": "53111"
      }
    },
  "links": [
    {
      "rel": "within",
      "href": "http://data.example.com/
               collections/parcels/items/
               DENW36AL100073O2FL",
      "title": "210",
      "type": "application/geo+json"
      }
    ]
}
```

As discussed above, it is not common to specify a schema for the properties and, therefore, currently no standardized encoding rule exists for representing an application schema in JSON Schema. Therefore, two comments on alternative approaches to encoding the building feature in GeoJSON are made here:

- Software capable of processing GeoJSON may be limited to content in the "properties" object that uses simple values only. "name", for example, has a simple value, but "use" or "address" have structured values (a JSON array or a JSON object). In order to create GeoJSON that all software should be able to handle, the structured values need to be "flattened". For example, as follows:

```
...
"use-1": "Museum",
"address.streetName": "Bonngasse",
"address.houseNumber": "20",
"address.city": "Bonn",
"address.postCode": "53111",
...
```

- JSON has no built-in capability to express links. GeoJSON does not specify any mechanism to represent links either. One approach to add links in a generic way was used in the example above. Another option would have been to include a "within" member in the "properties" object with the target URI as its value:

```
...
"within ": "http://example.com/data/
            parcels/DENW36AL10007302FL",
...
```

The same building encoding as GML, according to the standard GML encoding rule, is shown below:

```
<app:Building gml:id="DENW36AL10006wORBL">
 <app:footprint>
  <gml:Polygon srsName="http://www.opengis
                         .net/def/crs/OGC/
                         1.3/CRS84">
   <gml:exterior>
    <gml:LinearRing>
     <gml:posList> 6.77145389740199
                  51.513055921535326
                   6.771456210102157
                  51.51303866250359
                   6.77167501551265
                  51.51303808155525
                   6.771674222426856
                  51.51303897479996
                   6.771675139403735
                  51.513089282405375
                   6.771453749979931
                  51.51308846525745
                   6.77145389740199
                  51.513055921535326
     </gml:posList>
    </gml:LinearRing>
   </gml:exterior>
  </gml:Polygon>
 </app:footprint>
 <app:name>Beethovenhaus</app:name>
 <app:use>Museum</app:use>
 <app:address>
  <app:Address>
   <app:streetName>
    Bonngasse
   </app:streetName>
   <app:houseNumber>20<app:houseNumber>
   <app:city>Bonn</app:city>
   <app:postCode>53111</app:postCode>
   <app:country>Germany</app:country>
  </app:Address>
 </app:address>
 <app:within
    xlink:href="http://example.com/data/
                parcels/DENW36AL10007302FL"
   xlink:title="210"/>
</app:Building>
```

Options for the encoding of the GML application schema to XML elements
In the case of GML, the encoding rule specifies in detail how an application schema in UML is converted into a GML application schema in XML Schema. Since the conversion rules map aspects from the application schema into XML Schema, this implicitly specifies how instances are mapped, too. For example, the property "within" is converted to an XML element "app:within" under the XML element "app:Building".

However, the data output structure schema may still include some encoding options, and in this case, the encoding service may select the most appropriate option. A typical example in GML is the option of whether the cadastral parcel is referenced from the building using an Xlink [16] (as was done in the example above) or whether the XML element representing the parcel is embedded inline as a child element of the "within" property element.

Schema and Instance Level

Figure 4.3 also highlights that an encoding rule specifies conversion rules at both the schema and the instance levels.

- At the schema level, the conversion rules define a mapping for each of the concepts defined in the application schema to corresponding concepts in the data structure schema.
- At the instance level, the conversion rules define a mapping for each of the instances in the input data structure to corresponding instances in the output data structure. The instance conversion rules may be deduced from the schema conversion rules.

Examples of conversion rules on schema and instance levels

A conversion rule on the schema level specifies how an element of the application schema is represented in the Data Structure Schema. For example, the property "footprint" of the feature type "Building" is mapped in GML to the declaration of an XML element:

```
<xs:element name="footprint"
  type="gml:SurfacePropertyType"/>
```

In a GeoJSON encoding rule, it would be mapped to the "geometry" member of the GeoJSON feature schema (http://geojson.org/schema/Feature.json). Note that since GeoJSON uses a pre-defined member for the geometry; the term "footprint" is not used.

On the instance level, the footprint of the building is mapped to a valid gml:Polygon element in GML or to a valid Polygon geometry in GeoJSON, as shown in the instance examples above.

Other General Encoding Requirements

For an unambiguous encoding, additional information must be specified as part of an encoding rule:

- The order of bits within each byte, and bytes within a word, if applicable.
- The character repertoire and encoding; encoding rules typically use the Unicode standards and its character-encoding schemes [17].
- Any exchange metadata that is required to process the encoded data structure.
- The conventions for the identification of the dataset and the objects within the dataset.

4.1.3 The Role of Multiple Encoding Rules

There is no fixed set of encoding requirements that would allow the specification of only a single encoding rule for each data structure schema language. Different use cases and different requirements result in different encoding rules; for example, currently, the ISO 19100 series of international standards specifies two XML-based encoding rules:

- GML (ISO 19136) specifies an XML-based encoding rule for ISO 19109-conformant application schemas that can be represented using a restricted profile of UML that allows for a conversion to XML schema. The encoding rule has mainly been developed for the purpose of application schemas specifying feature types and their properties. The encoding rule uses XML schema for the output data structure schema.
- ISO/TS 19139-1 specifies an XML-based encoding rule for conceptual schemas specifying types that describe geographic resources, e.g., metadata according to ISO 19115-1 *Metadata* and feature catalogues according to ISO 19110 *Methodology for feature cataloguing*. The encoding rule supports the UML profile as used in the UML models commonly used in the standards developed by ISO/TC 211. The encoding rule uses XML Schema for the output data structure schema.

This means that different encoding rules may be required by the information community for different purposes. Examples of relevant requirements include

- Support for an existing output data structure schema that is supported by existing applications and where the cost of changes that break backwards compatibility would outweigh the benefits.
- Support for a different profile of a conceptual schema language that is not covered by an existing encoding rule. Examples are UML profiles supporting multiple inheritance, which is not supported by the encoding rules of GML and ISO/TS 19139.
- Support for a different output data structure schema language or a new version or variant of such a language. Examples could be encoding rules targeting GeoJSON, GeoPackages, or Shapefiles, or that use JSON schema, SQL DDL, or the web ontology language (OWL) to encode the data structure.
- Support for specific conversions to optimize the use of the capabilities of the output data structure schema language. For example, since coordinate data is a significant part of geographic information, GeoJSON, GML, and SQL/MM [34] specify optimized encodings for geometries and coordinate data, as discussed above.
- Support for other requirements related to the output data structure schema that are established in a community.

4.2 Sample Encoding Rules

As mentioned above, GML specifies an XML-based encoding rule using XML Schema for the output data structure schema.

4.2.1 GML Encoding Rule

The GML Encoding Rule supports application schemas that can be represented using a restricted profile of UML that al-

Table 4.1 GML model representation in terms of the concepts of an application schema in UML

	Representation in UML	Representation in GML instances
Features	By objects, where the name of the feature type is used as the name of the object class	By XML elements, where the name of the feature type is used as the name of the element
Feature properties	By association roles with feature type classes and attributes of feature type classes, where the property semantics are given by the association role name or attribute name	By child elements (known as property elements) of feature elements, where the property semantics are given by the property element name
Property value type	By the class of the association target or by the data type of the attribute	In the case of properties with complex values, by the name of the object element contained within the property element; in the case of a property with simple values, by the type of the literal value containing no embedded XML markup

lows for conversion to XML Schema (see annex E.2.1 in the GML document [3]); for example, the current version 3.2 does not support multiple inheritance [2], and association classes are only supported in an extension in version 3.3 [3].

On the other hand, to support the conversion to GML/XML, additional stereotypes such as «featureType» to identify classes that represent features and additional tagged values such as *targetNamespace* to identify the XML namespace of the converted GML application schema are added to the UML profile.

The encoding rule was mainly developed for the purpose of application schemas specifying feature types and their properties.

A feature is encoded as an XML element with the name of the feature type. Other identifiable objects are encoded as XML elements with the name of their type, too.

Each feature attribute and feature association role is a property of a feature. Feature properties are encoded in an XML element that is a child element of the feature or object.

The property semantics, which is indicated by the name of the element representing the property, is distinguished from the property value, which is given by the content of the property element. A property element may contain its value as content encoded inline or reference its value with a simple Xlink [16]. The value of a property may be a simple value or an object – represented by an element.

When recorded inline, the value of a simple property is recorded as a literal value with no embedded markup (text), while if the value is complex, it appears as a subtree using XML markup (i. e., an XML element with substructure).

As a result, the GML model has a straightforward representation in terms of the concepts of an application schema in UML, which can be summarized approximately and briefly as shown in Table 4.1.

The result is a layered XML document, in which XML elements corresponding to features, objects, or values occur interleaved with XML elements corresponding to the properties that relate them. The function of a feature, object, or value in context can always be determined by inspecting the name of the property element that directly contains it or carries the reference to it.

This encoding pattern is sometimes referred to as the *object–property model*. While in some cases, this encoding pattern adds extra levels of elements in instance documents, it also provides significant benefits. It helps to make a GML instance document understandable on its own, provides a predictable structure, independent of the schema language, W3C XML Schema [13].

Encoding values inline or remote resources in GeoJSON and GML

The value of the properties *name, address*, and *footprint* are encoded inline. In the case of *name*, the value is simple – a string – and for *footprint*, it is another object – a polygon geometry. The value of *within* is obtained by following the hyperlink (an "xlink:href" XML attribute in GML or an "href" member in a link object in the JSON).

4.2.2 ISO/TS 19139-1 Encoding Rule

ISO/TS 19139-1 specifies another XML-based encoding rule. Its scope is conceptual schemas specifying types that describe geographic resources such as metadata according to ISO 19115-1 or feature catalogues according to ISO 19110.

The encoding rule uses XML schema for the output data structure schema.

XML instances created according to the encoding rule follow a similar pattern to GML instances. In particular, they follows the *object–property–model* described in Sect. 4.2.1, too.

However, there are differences (between the encoding rules of ISO/TS 19139-1 and GML) in the details; for example, a property value that is a simple type (e.g., a string or a number) is not represented in GML by a literal value, while in an ISO/TS 19139-1 encoding the literal value is embedded in an XML element with the name of the simple type, e.g., CharacterString.

Different schema conversion rules in GML and in ISO/TS 19139-1

Where the GML encoding rule converts the string-valued property *owner* to

```
<app:name>Beethovenhaus</app:name>
```

the ISO/TS 19139 encoding rule would convert this property to

```
<app:name>
 <gco:CharacterString>Beethovenhaus
 </gco:CharacterString>
</app:name>
```

4.3 Commonly Used Formats to Encode Geographic Information

This section provides a brief overview of data formats that are commonly used to encode geographic information.

The W3C/OGC Spatial Data on the Web Best Practices include an annex that discusses the applicability of common formats on the web. Key formats are the following:

- GeoJSON [6] is a data format based on JavaScript object notation (JSON), extending it to support features, feature collections, and geometries. GeoJSON has limitations but is easy to use and well supported by software.
  JSON [18] was originally derived from JavaScript – hence the name – but is a text-based, standardized data format that today is supported in most programming languages and environments. JSON became popular in web development due to being a native encoding for structured data in JavaScript. JavaScript [19] is the most important scripting language of the web.
- GML is the geography markup language developed by the Open Geospatial Consortium (OGC), also available as an ISO standard. GML is an XML [20] grammar to encode geographic features and their properties designed for use on the web. It offers built-in support for a wide range of types commonly used in the context of geographic information, including spatial and temporal geometry and topology, coordinate reference systems, grids, and coverages. Due to its richness, it is more complex to use than, for example, GeoJSON.
- GML application schemas such as CityGML, an OGC standard, are specified and used by communities to define a standardized XML encoding for a particular application schema, in this case 3D city models.
- GeoRSS [21] is a lightweight extension to Atom [22] and Really Simple Syndication (RSS) [23] to enable geographically tagged web feeds. Web feeds allow software programs to check for updates published on a website, and Atom and RSS are the main specifications used to implement web feeds. Two variants of GeoRSS exist, depending on the capabilities required: GeoRSS-Simple supports basic geometries and WGS 84 latitude/longitude coordinates, and GeoRSS-GML supports additional capabilities, e.g., other coordinate reference systems.
- GeoSPARQL [7] is an OGC standard and defines extensions for the W3C SPARQL standard [24] to add a set of terms and functions for modeling and querying spatial information in the Resource Description Framework (RDF) [25]. RDF is a family of W3C standards used to model information on the web. SPARQL is an RDF query language. GeoSPARQL adds spatial capabilities, and geometries are encoded by using WKT (see [33, 34]) or GML (see earlier).
- The schema.org vocabulary [26] was designed for annotating HTML web pages with machine-readable metadata, see Example 8 in [1]. The vocabulary supports location information, including geometries, for example, the bounding box of a dataset or the coordinates of a place. HTML pages annotated with such information can be understood by software agents parsing HTML pages, including crawlers from search engines.

The following encodings have been designed in particular to support visualization, including on mobile devices or in web browsers

- KML [27] is another OGC standard, originally developed by Keyhole, Inc. KML is an XML grammar focused on geographic visualization, including annotation of maps and images. Geographic visualization includes not only the presentation of graphical data but also the control of the user's navigation in the sense of where to go and where to look.
- i3s [28] and 3D Tiles [29] are OGC community standards that support streaming and rendering large amounts of 3D data. They organize the feature data so that exactly the data that is needed for display from a certain viewpoint is streamed to the client within the appropriate level of detail, depending on the distance from the viewpoint, etc.
- MBTiles [30] and Mapbox Vector Tiles [31] serve the same purpose but are optimized for a 2-D map view. Feature data is organized in pyramids of tiles with different levels of detail, depending on the zoom level.

Another common form of organizing datasets are tables. Some typical encodings as tabular data are

- Shapefiles [5] are still one of the most commonly used and supported data formats for geographic information. Every geographic feature has a geometry (point, line, or polygon) plus attribute information. Attributes are simple and are encoded using the dBASE file format.
- GeoPackage [18] is an OGC standard to support the exchange and direct use of vector geospatial features and/or tile matrix sets of earth images and raster maps at various scales in an SQLite database [32]. SQLite is available as a library and is popular for embedded storage in applications, for example, in web browsers or smartphones. Direct use means the ability to access and update data in a "native" storage format without intermediate format translations. GeoPackages are particularly useful on mobile devices and were designed to communications environments with limited connectivity and bandwidth.
- The SQL standard has been extended with two encodings (well-known text (WKT), and well-known binary (WKB)) to support geometries as native values in SQL. As the names suggest, WKT is a text-based encoding, and WKB is a binary encoding for geometries. Originally, the SQL extensions were specified in the OGC standard simple feature access [33], but the latest definition are part of ISO/IEC 13249-3 [34].

As described in Chap. 1, coverages are another commonly used representation of spatial data, where one or more properties vary their values depending on the location in space and time

- GeoTIFF [35] is a standard that allows georeferencing information to be embedded within a tagged image file format (TIFF) file [36], a file format for storing images.
- Cloud Optimized GeoTIFF (COG) is variant that is "a regular GeoTIFF file, aimed at being hosted on a HTTP file server, with an internal organization that enables more efficient workflows on the cloud. It does this by leveraging the ability of clients issuing HTTP GET range requests to ask for just the parts of a file they need [37]."
- Joint Photographic Experts Group (JPEG) 2000 is a wavelet-based image compression standard and coding system; it is a successor to the widely used JPEG format. The main advantage offered by JPEG 2000 is significant flexibility. GML in JPEG 2000 for geographic imagery encoding (GMLJP2), an OGC standard [38], utilizes the additional flexibility provided by GML and standardizes how GML documents can be stored as part of a JPEG 2000 code stream to provide information about georeferencing, coordinate reference systems, etc.
- Network Common Data Form (netCDF) [39] is an interface for array-oriented data access and a library that provides an implementation of the interface. The netCDF library also defines a machine-independent format for representing scientific data, including spatial and temporal aspects.
- The coverage implementation schema (CIS) [40] is an OGC and ISO standard for representing all kinds of coverage data. It extends the GML coverage model. CIS is mainly used in the OGC Web Coverage Service standard.

This is just a small selection of existing data formats; many others exist and are in active use. In addition, new JSON-based formats are emerging (for example, CovJSON or CityJSON).

In practice, geographic information will often be available in processed form and encoded in a format that is supported by other applications that are not geographically enabled. Examples of such formats include HTML without schema.org annotations, comma-separated values (CSV), scalable vector graphics (SVG), extensible 3D (X3D), portable document format (PDF), and others.

4

References

1. W3C/OGC: Spatial Data on the Web Best Practices, W3C Working Group Note (2017). https://www.w3.org/TR/sdw-bp/ (Created 28.09.)
2. OGC: Geography Markup Language (GML): v.3.2, OGC Document 07-036r1 (2018). http://www.opengeospatial.org/standards/gml. same as ISO 19136-1:2019 Geographic Information – Geography Markup Language (ISO, Geneva 2019), to be published
3. OGC: Geography Markup Language (GML): v.3.3, OGC Document 10-129r1 (2012). http://www.opengeospatial.org/standards/gml. same as ISO 19136-2:2015: Geographic Information – Geography Markup Language, Part 2: Extended schemas and encoding rules (ISO, Geneva 2015)
4. OGC: GeoPackage, v.1.2 (2017). https://www.opengeospatial.org/standards/geopackage
5. Esri: ESRI Shapefile Technical Description (1998). https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf
6. IETF: The GeoJSON Format, RFC 7946, Internet Engineering Task Force (IETF). https://www.rfc-editor.org/rfc/rfc7946.txt
7. OGC: GeoSPARQL, v.1.0 (2012). https://www.opengeospatial.org/standards/geosparql
8. ISO 19118: Geographic Information – Encoding. ISO, Geneva (2011)
9. ISO/TS 19139-1: Geographic Information – Metadata – XML Schema Implementation. ISO, Geneva (2019)
10. IANA: Media type application/geo+json (2016). https://www.iana.org/assignments/media-types/application/geo+json
11. IANA: Media type application/gml+xml (2017). https://www.iana.org/assignments/media-types/application/gml+xml
12. IANA: Media type application/geopackage+sqlite3 (2018). https://www.iana.org/assignments/media-types/application/geopackage+sqlite3
13. W3C: XML Schema Part 0: Primer. W3C Recommendation (2004). https://www.w3.org/TR/xmlschema-0/ (Created 28.10.)
14. IETF: JSON Schema: A Media Type for Describing JSON Documents (2017). https://json-schema.org/ (Created 19.11.). Draft
15. ISO/IEC 9075-2:2016: Information Technology – Database Languages – SQL – Part 2: Foundation (SQL/Foundation). ISO, Geneva (2016)

16. W3C: XML linking language (XLink), v.1.0, W3C Recommendation (2001). https://www.w3.org/TR/xlink/
17. The Unicode Consortium: (2010). http://unicode.org/
18. Internet Engineering Task Force (IETF): The JavaScript Object Notation (JSON) Data Interchange Format, RFC 8259. https://www.rfc-editor.org/rfc/rfc8259.txt
19. Ecma International: ECMAScript® 2018 Language Specification, ECMA-262 (2018). https://www.ecma-international.org/publications/standards/Ecma-262.htm
20. W3C: XML, Extensible Markup Language (XML) 1.0, 5th edn., W3C Recommendation (2008). https://www.w3.org/TR/xml/ (Created 26.11.)
21. OGC: GeoRSS Encoding Standard, v.1.0 (2017). http://www.opengeospatial.org/standards/georss
22. Internet Engineering Task Force (IETF): The Atom Syndication Format, RFC 4287. https://www.rfc-editor.org/rfc/rfc4287.txt
23. Wikipedia: RSS. https://en.wikipedia.org/wiki/RSS
24. W3C: SPARQL 1.1 Overview, W3C Recommendation (2013). https://www.w3.org/TR/sparql11-overview/ (Created 21.03.)
25. W3C: RDF 1.1 Primer, W3C Working Group Note (2014). https://www.w3.org/TR/rdf11-primer/ (Created 24.06.)
26. schema.org. https://schema.org/
27. OGC: KML, v.2.3 (2015). http://www.opengeospatial.org/standards/kml
28. OGC: Indexed 3D Scene Layers (i3s), v.1.0 (2017). https://www.opengeospatial.org/standards/i3s
29. OGC: 3D Tiles, v.1.0 (2019). https://www.opengeospatial.org/standards/3DTiles
30. Mapbox: MBTiles Specification, v.1.3. https://github.com/mapbox/mbtiles-spec
31. Mapbox: Mapbox Vector Tile Specification, v.2.1. https://github.com/mapbox/vector-tile-spec
32. SQLite. https://www.sqlite.org/
33. OGC: Simple Feature Access – Part 2: SQL Option, v.1.2.1 (2010). https://www.opengeospatial.org/standards/sfs
34. ISO/IEC 13249-3:2016: Information Technology – Database Languages – SQL Multimedia and Application Packages – Part 3: Spatial. ISO, Geneva (2016)
35. GeoTIFF: Format Specification, v.1.8.2. http://geotiff.maptools.org/spec/geotiffhome.html
36. Adobe: TIFF, Revision 6.0 (1992). https://www.adobe.io/content/dam/udp/en/open/standards/tiff/TIFF6.pdf
37. Cloud Optimized GeoTIFF. https://www.cogeo.org/
38. OGC: GML in JPEG 2000 for Geographic Imagery Encoding, v.2.1 (2018). http://www.opengeospatial.org/standards/gmljp2
39. OGC: netCDF, v.1.0. http://www.opengeospatial.org/standards/netcdf
40. OGC: Coverage Implementation Schema ("CIS"), v.1.1 (2017). http://www.opengeospatial.org/standards/wcs

Clemens Portele Clemens Portele holds a diploma in Physics from the University of Bonn, Germany. Since 1996 he has been a co-owner and Managing Director of interactive instruments, where the current focus of his work is spatial data management based on open standards, with an emphasis on bringing spatial data and Web technologies closer together.

5 Big Data Analytics

Erik Hoel

Contents

Abstract

Big data analytics, in the context of geospatial data, employs distributed computing using advanced tools that support spatiotemporal analysis, spatial statistics, and machine learning algorithms and techniques (e. g., classification, clustering, and prediction) on very large spatiotemporal datasets to visualize, detect patterns, gain deeper understandings, and answer questions. In this chapter, the key definitions, domain-specific problems, analysis concepts, current technologies and tools, and remaining challenges are discussed.

Keywords

analytics · big data · cloud platforms · distributed computing

E. Hoel (✉)
Research and Development, Esri
Redlands, CA, USA
e-mail: ehoel@esri.com

5.1 Overview

Big data analytics involves analyzing large volumes of varied data, or big data, to identify and understand patterns, correlations, and trends that are ordinarily invisible due to the volumes involved in order to allow users and organizations to make better decisions. These analytics, in the context of geospatial data, commonly involve spatial processing, sophisticated spatial statistical algorithms, and predictive modeling. Big data can be obtained from a wide variety of sources; this includes sensors (both stationary and moving), aerial and satellite imagery, lidar, videos, social networks, website activity, sales transaction records, and real-time stock trading transactions. Users and data scientists apply big data analytics to evaluate these large collections of data; data with volumes that traditional analytical systems are unable to accommodate [1]. This is particularly the case with unstructured or semistructured data (such data types are problematic with data warehouses, which often utilize relational database concepts and work with structured data).

To address these complex demands, many new analytic environments and technologies have been developed. This includes distributed processing infrastructures (e. g., Spark and Map Reduce [2–4]), distributed file stores, and NoSQL databases [5–9]. Many of these technologies are available in open-source software frameworks such as Apache Hadoop [10] that can be used to process huge data sets with clustered systems.

When working with big data, there is a collection of objectives that users have when performing big data analytics [11, 12]; these include:

- *Discovering value from big data.* Visualizing and analyzing big data in a way that reveals patterns, trends, and relationships that traditional reports and spatial processing do not. Data may exist in many disparate places, streams, or web logs.
- *Exploiting streaming data.* Filter and convert raw streaming data from various sources that contains geographical

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_5

elements into geographic layers of information. The geographical layers can then be used to create new, more useful maps and dashboards for decision-making.

- *Exposing geographic patterns.* Using maps and visualization to see the story behind the data. Examples of identifying geographical patterns include retailers seeing where promotions are most effective and where the competition is, banks understanding why loans are defaulting and where there is an underserved market, and climate-change scientists determining the impact of shifting weather patterns.
- *Finding spatial relationships.* Seeing spatially enabled big data on a map allows you to answer questions and ask new ones. Where are disease outbreaks occurring? Where is insurance risk greatest given recently updated population shifts? Geographic thinking adds a new dimension to big data problem solving and helps you make sense of big data.
- *Performing predictive modeling.* Predictive modeling using spatially enabled big data helps you develop strategies from if/then scenarios. Governments can use it to design disaster response plans. Natural resource managers can analyze recovery of wetlands after a disaster. Health service organizations can identify the spread of disease and ways to contain it.

What Differentiates Spatial Big Data

Spatial big data is differentiated from standard (nonspatial) big data by the presence of spatial relationships, geostatistical correlations, and spatial semantic relations (this can be generalized to include the temporal domain [13]). Spatial big data offers additional challenges beyond what is encountered with more traditional big data. Spatial big data is characterized by the following [14]:

- *Volume.* The quantity of data. Spatial big data also include global satellite imagery, mobile sensors (smart phones, GPS trackers, and fitness monitors), and georeferenced digital camera imagery.
- *Variety.* Spatial data is composed of 2-D or 3-D vector or raster imagery. Spatial data is more complex and subsumes the types found with conventional big data.
- *Velocity.* Velocity of spatial data is significant given the rapid collection of satellite imagery in addition to mobile sensors.
- *Veracity.* For vector data (points, lines, and polygons), the quality and accuracy varies. Quality is dependent upon whether the points have been GPS determined, determined by unknown origins, or manually. Resolution and projection issues can also alter veracity. For geocoded points, there may be errors in the address tables and in the point location algorithms associated with addresses. For raster data, veracity depends on the accuracy of recording instruments in satellites or aerial devices, and on timeliness.
- *Value.* For real-time spatial big data, decisions can be enhanced through visualization of dynamic change in such spatial phenomena as climate, traffic, social-media-based attitudes, and massive inventory locations. Exploration of data trends can include spatial proximities and relationships.

  Once spatial big data are structured, formal spatial analytics can be applied, such as spatial autocorrelation, overlays, buffering, spatial cluster techniques, and location quotients.

5.2 Definitions

The following terms are referenced in this chapter and are included here to facilitate a more rapid understanding of the general concepts discussed later in this chapter.

- *Amazon Web Services (AWS):* A secure, on-demand, cloud computing platform where users pay for the computing resources that they consume (e. g., computing, database storage, and content delivery).
- *Artificial intelligence:* Computer systems or machines that are able to perform tasks and mimic behavior that normally requires human intelligence, such as visual perception, speech recognition, and language translation.
- *Big data as a service (BDaaS):* Cloud-based hardware and software services that support the analysis of large or complex data sets. These services can provide data, analytical tools, event-driven processing, visualization, and management capabilities.
- *Cloudera:* A software company that provides a software platform that can run either in the cloud or on-prem, supporting data warehousing, machine learning, and big data analytics. The company is a major contributor to the Apache Hadoop platform (e. g., Avro, HBase, Hive, and Spark).
- *Computer vision:* A scientific discipline that focusses on the acquisition, extraction, analysis, and understanding of information obtained from either single or multidimensional images or video data.
- *Data as a service (DaaS):* Built on top of software as a service, data is provided to users on demand for further processing and analysis. The centralization of the data enables higher-quality curated data at a lower cost to the client.
- *Databricks:* A company that provides a cloud-based platform for working with Apache Spark. Databricks traces it origins to the AMPLab project at Berkeley, which evolved into an open-source distributed computing framework for working with big data.
- *Data mining:* The process of discovering and extracting hidden patterns and knowledge found in big data us-

ing methods and techniques commonly associated with database management, machine learning, and statistics.

- *Deep learning:* A subfield of machine learning that focusses on algorithms and computational architectures that mimic the structure of the brain (commonly termed artificial neural networks). Recent advances in large-scale distributed processing have enabled the development and use of very large neural networks.
- *Elastic Compute Cloud (EC2):* Infrastructure within Amazon Web Services (AWS) that provides scalable computing capacity; clients can develop, deploy, and run their own applications. EC2 is elastic and allows clients to scale their compute and storage up or down as necessary.
- *Hadoop:* Apache Hadoop is an open-source framework and set of software modules that enable users to solve problems on big datasets using a distributed cluster of hardware resources. This includes distributed data storage and computation using the MapReduce programming model. Hadoop was originally inspired by Google's work in the distributed processing domain.
- *HDFS (Hadoop Distributed File System):* A distributed and scalable file system and data store that is part of Apache Hadoop. HDFS stores big data files across a cluster of machines and supports high reliability by replication of the data across different nodes in the cluster.
- *Hive:* Data warehouse software module in Apache Hadoop that facilitates querying and analyzing big data stored in HDFS in a distributed and replicated manner using an SQL-like language termed HiveQL.
- *IBM Cloud:* A set of cloud computing capabilities and services that provides capabilities including software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS).
- *Infrastructure as a service (IaaS):* A type of cloud computing infrastructure that virtualizes computing resources, storage, data partitioning, scaling, and networking. Unlike software as a service (SaaS) or platform as a service (PaaS), IaaS clients must maintain the applications, data, middleware, and operating system.
- *Machine Learning:* A subset of artificial intelligence where software systems can automatically learn and improve without any explicit programming, relying upon statistical methods for pattern detection and inference. Machine learning software creates statistical models using sample data in order to make decisions or predictions.
- *MapReduce:* A programming model, originally developed at Google, that is often used when processing big datasets in a distributed manner. MapReduce programs contain a map procedure where data can be sorted and filtered, and a reduce procedure where summary operations are performed. MapReduce systems, such as Apache MapReduce, are responsible for managing communications and data transfer among the collection of distributed processing nodes.
- *Microsoft Azure:* A cloud computing service from Microsoft for creating, deploying, and managing applications using data centers managed by Microsoft. Hundreds of services are available, which provide functionality related to compute, data management, messaging, mobile, and storage capabilities.
- *Natural language processing (NLP):* A portion of artificial intelligence that focusses on enabling computer to understand and communicate (including language translation) through human language, both written and spoken.
- *NoSQL data stores:* A non-SQL or nonrelational database that provides a mechanism for storage and retrieval of data. NoSQL data stores often trade consistency in favor of availability, speed, horizontal scalability, and partitionability.
- *Oracle Cloud:* A collection of cloud computing services from Oracle providing servers, storage, network, applications, and services using Oracle-managed data centers. The Oracle Cloud provides software as a service (SaaS), platform as a service (PaaS), infrastructure as a service (IaaS), and data as a service (DaaS).
- *Pig:* Apache Pig is used to develop programs for analyzing big datasets that run on Apache Hadoop using a high-level language (Pig Latin). Pig can be used to develop functionality that runs as MapReduce, Tez, or Spark jobs.
- *Platform as a service (PaaS):* A category of Cloud computing service that allows clients to develop, deploy, run, and manage applications without needing to build or maintaining the Cloud computing infrastructure. Unlike software as a service (SaaS), the client is responsible for maintaining the applications and data.
- *Predictive analytics:* A group of statistical and machine learning algorithms that are used to predict the likelihood of future or other unknown events based upon existing historical data.
- *Real-time data processing:* A collection of software and hardware that processes data on the fly and is subjected to a constraint where responses must be provided within a short interval of time (e. g., fractions of a second), independent of system or event data load.
- *Redshift:* Amazon Redshift is a column-oriented, fully managed, big data warehouse. Redshift is similar to other columnar NoSQL databases, as it is intended to scale out with distributed clusters of low-cost hardware.
- *Simple Storage Service (S3):* Amazon Simple Storage Service is an object storage service offered by Amazon Web Services (AWS); it is intended to store any type of data (objects) that can later be used for big data analytic processing.
- *Software as a service (SaaS):* A category of cloud computing service that allows clients to license applications, web-based software, on-demand software and hosted soft-

ware. The delivery model is on a subscription basis and is centrally hosted. Differing from platform as a service (PaaS), SaaS does not require client to manage either data or software.

- *Spark:* An analytic engine and cluster-computing framework, part of Apache Hadoop, which supports applications that run across a distributed cluster. Originally developed at Berkeley in 2009, it provides a framework for programming clusters of machines with data parallelism.
- *Speech recognition:* A collection of methodologies and techniques that enables the recognition and transformation of spoken language into text for further computational processing.
- *Storm:* Apache Storm is a real-time, distributed, high-volume, stream processing framework for big data. It is part of the Apache Hadoop open-source framework.
- *Stream processing:* A computer programming paradigm (similar to dataflow programming), where given a sequence of data (a stream), a series of pipelined operations (or kernel functions), is applied to each element in the stream.

5.3 Example Problems

There are a significant number of industries and application domains that benefit from spatiotemporal big data analytics [15]. As the sheer number of processes and technologies that collect spatial data grows, the ubiquity and significance of the data has grown. Spatial big data analytics has wide applicability and value across numerous domains; a few of these are the following:

Agriculture
Farmers can use spatial big data analytics to detect and analyze patterns in weather data, correlated with historical crop yields, surface topography, and soil characteristics. This helps farmers determine the best times, seed varieties, and places to plant crops in order to maximize yields. In addition, the distribution of fertilizer can also be optimized based upon historical information. Tractor and heavy equipment movement can also be tracked via GPS and incorporated into the logistic optimization analytics, as well as identifying the area of usable and productive land within a field.

Commercial retailers
Commercial retailers have always used local shopping patterns and demographics to drive marketing strategies and site selection. However, retailers can now use spatial big data analytics to analyze the locations and characteristics of customers along with social media conversations and browsing behavior in order to better understand customers' needs. Retailers can essentially build a richer and more useful understanding and relationship with their customer base. New store site selection on regional or national levels can be optimized based on the locations of customers, competitors, and other nontraditional data.

Connected Cars
Developers of systems for connected cars and autonomous vehicles can use spatial big data analytics to provide accurate situational awareness to drivers and vehicles about their surrounding environment. Systems can apply analytics capabilities such as road snapping, predictive road snapping, change detection of objects sensed by the vehicle but not on the map, and accident prediction. This is all under the topic of improved vehicle reliability and passenger safety.

Environmental
Environmental organizations can employ spatial big data analytics to answer a number of important questions including whether there are spatiotemporal correlations between species observations – this can be done by geographic area or species.

Financial Services
In the financial services/insurance industry, spatial big data analytics are used to overlay weather data with claim data to assist companies detect possible instances of fraud. In other contexts, nontraditional data sources like satellite imagery are combined with traditional topographic data sources to identify the potential risk of offering flood insurance. Insurers can also assess spatial relationships between their insurance portfolios and past hazards to balance risk exposure. Finally, banks can use spatiotemporal historical transaction data to help them detect evidence of fraud.

Government Agencies
National and regional government agencies would like to use spatial big data analytics to process and overlay nation-wide datasets containing land use, parcels, planning information, geological informational, and environmental data in order to create information products that can be used by analysts, scientists, and policy makers to make better policy decisions (Fig. 5.1).

Health Care
Public health agencies can use spatial big data analytics to see how far patients are from health facilities, helping them evaluate access to care. Hospital networks can determine the density of hospitals in certain areas to identify gaps and opportunities. They can also measure the prevalence of certain habits and illnesses in the community using demographic data. Public health agencies can also utilize tracking data to perform contact tracing of infected individuals to identify

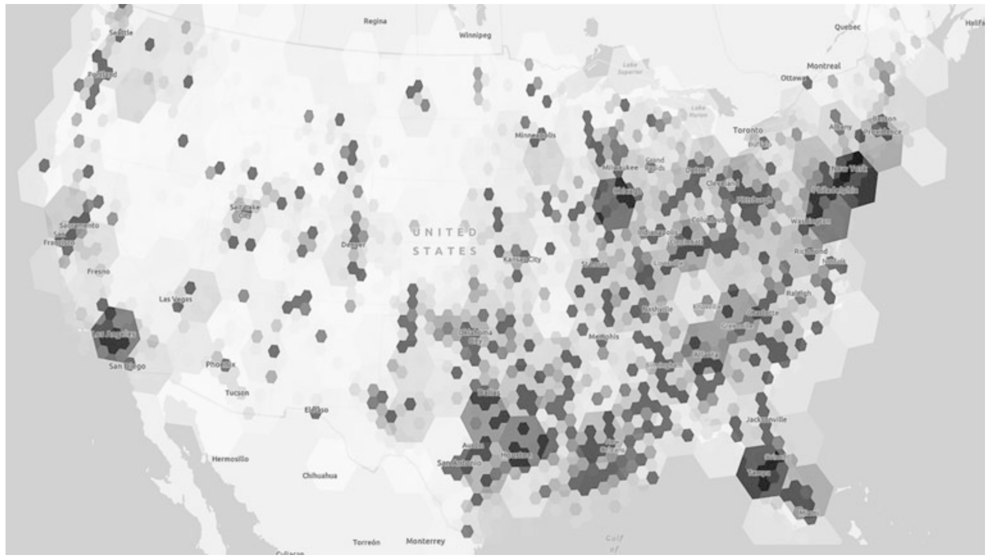


Fig. 5.1 Leveraging feature binning technology to see geographic trends between industrial emission activity in 2014 (small hexes) as reported in the EPA Toxic Release Inventory and total US electrical generation by load (large hexes) in 2018 as published by the Homeland Infrastructure Foundation-Level Data. (Sources: Esri, Garmin, FAO, NOAA, USGS, © OpenStreetMap contributors, and the GIS User Community)

who they have been in contact with in the past. The contact information can then be utilized to help reduce the infections in the general population. Proximity tracing is a variant where to contact is specified using proximity-based filtering criteria (e. g., spatial and temporal range) in order to identify potential contact events.

Marketing
Geospatial big data analytics is frequently used in corporate marketing for prospect and customer segmentation. Data from body sensors (e. g., smartphones, smart watches, fitness monitors) can be used to segment the customer base according to physical activity or behavioral patterns and deliver advertising in a targeted manner. Companies also want to be able to identify where their customers are in relation to their competitors' customers. This allows them to identify areas where they are losing the market and help determine where they need to focus their marketing efforts.

Mining
Mining companies can apply spatial big data analytics to perform complex vehicle tracking analysis to find ways to better manage equipment moves. For example, they can analyze patterns of equipment locations when braking and they can review shock absorption, RPM changes, and other telematics information. They can also analyze geochemical sample results.

Petroleum
Spatial big data analytics enable petroleum companies to identify suitable areas for exploration based upon historical production, geographic composition, and competitor activity (including leasing activity). Spatial big data analytics can also be used to review historical production data to assess reservoir production over time. Vehicle tracking data can be analyzed to determine time spent on both commercial and noncommercial roads. They can also review vessel tracks over offshore blocks using AIS vessel tracking information.

Retail
Retailers can use spatial big data analytics to model retail networks and help them select the best sites to optimize their store network. Analytic results can be used to create customer profile maps, allowing retailers to better understand customer behavior and the factors that influence their behavior. Retailers also want to spatially analyze the types of products that consumers are buying based upon seasonal and weather-related stimuli. This often incorporates promotions and sale activity. The spatiotemporal analysis can extend to a very fine-grained level – e. g., hourly sales activity on Black Friday.

Telecommunications
Telecommunications companies can use spatial big data analytics to review spatial trends in bandwidth usage over time to help plan new network deployments. They can analyze spatial patterns in consumer habits, spending patterns, demographics, and service purchases to improve marketing, define new products, and help plan network expansions. Customer service departments can correlate network problems and trouble tickets with customer complaints or cancellations to determine where and when service issues have led to customer dissatisfaction. Call detail records can be used to identify areas where cellular service is problematic (quality, speed, coverage), both temporally and spatially.

Transportation
With spatial big data analytics, commercial delivery companies can reconstruct vehicle routes from millions of individual position reports to check for routing inefficiencies and identify incidents of unsafe speeding and braking. This level of

visibility into past trips helps them develop strategies to improve efficiency and safety. Transportation planners can also use spatial big data analytics to aggregate, visualize, and analyze historical crash data for a metropolitan area, helping them identify unsafe road conditions. State and regional transportation agencies can analyze and model traffic slowdowns and congestion in order to optimize future road construction and rapid transit planning activities. City mobility planning (encompassing buses, ride sharing, and public bike systems) makes heavy use of spatio-temporal big data analytics in optimizing route planning and resource deployments in order to maximize throughputs and minimize congestion delays.

Utilities

Geospatial big data analytics is used by utility companies to summarize and analyze customer usage patterns across a service area. They can assess customer usage through time and correlate usage to weather patterns, helping them anticipate future demand. Utilities can also use spatial big data analytics to analyze SCADA (Supervisory Control And Data Acquisition – real-time remotely collected data), smart meter, and other sensor data to detect and quantify potential problems in the distribution network – such as when and where outages occurred, were they correlated with weather events, and how many customers were affected. They can use this information to prioritize maintenance activities and prevent or mitigate future problems. Public utility commissions consume raw energy data from utilities and prepare future forecasts of energy consumption. Energy efficiency can also be studied; what are the seasonal impacts and what can be done to guide consumers toward smarter energy usage.

5.4 Big Data Analysis Concepts

The types of analysis that may be performed against spatial big data often parallels that which is typically done with traditional spatial data [16]. However, when working with big data, it is oftentimes necessary to identify the key or most significant subsets of data in the larger collection. Once the interesting data is identified, further detailed analysis using the full breadth of spatio-temporal analysis tools and techniques can then be applied. This is particularly common when working with spatial big data that is obtained from sensors.

5.4.1 Summarizing Data

Summarizing data encompasses operation that calculate total counts, lengths, areas, and basic descriptive statistics of features and their attributes within areas or near other features. Common operations that summarize data include:

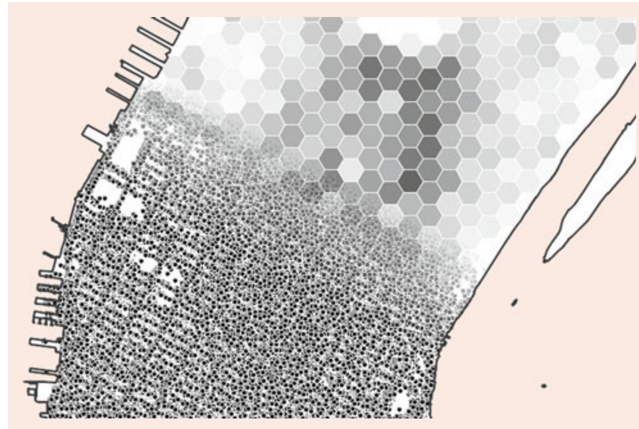


Fig. 5.2 Ridesharing pick up locations in Midtown Manhattan. In the southern portion of the figure, the raw data is shown. The northern region shows the data aggregated into 250 meter height hexagon cells

- *Aggregations* – aggregate points into polygon features or bins. At all locations where points exist, a polygon is returned with a count of points as well as optional statistics (Fig. 5.2).
- *Joins* – join (match) two datasets based upon their spatial, temporal, or attribute relationships. Spatial joins match features based upon their spatial relationships (e. g., overlapping, intersecting, within distance, etc.); temporal joins match features based upon their temporal relationships, and attributes joins match features based upon their attribute values.
- *Track reconstruction* – create line tracks from temporally enabled, moving point features (e. g., positions of cars, aircraft, ships, or animals).
- *Summarization* – overlay one dataset on another and calculate summary statistics representing these relationships. For example, one set of polygons may be overlayed on another dataset in order to summarize the number of polygons, their area, or attribute statistics.

5.4.2 Identify Locations

Location identification involves identify areas that meet a number of different specified criteria. The criteria can be based on attribute queries (for example, parcels that are vacant) and spatial queries (for example, within 1 km of a river). The areas that are found can be selected from existing features (such as existing land parcels), or new features can be created where all the requirements are met. Common operations that are used to identify locations include:

- *Incident detection* – detect all features that meet a specified criteria (e. g., lightning strikes exceeding a given intensity).

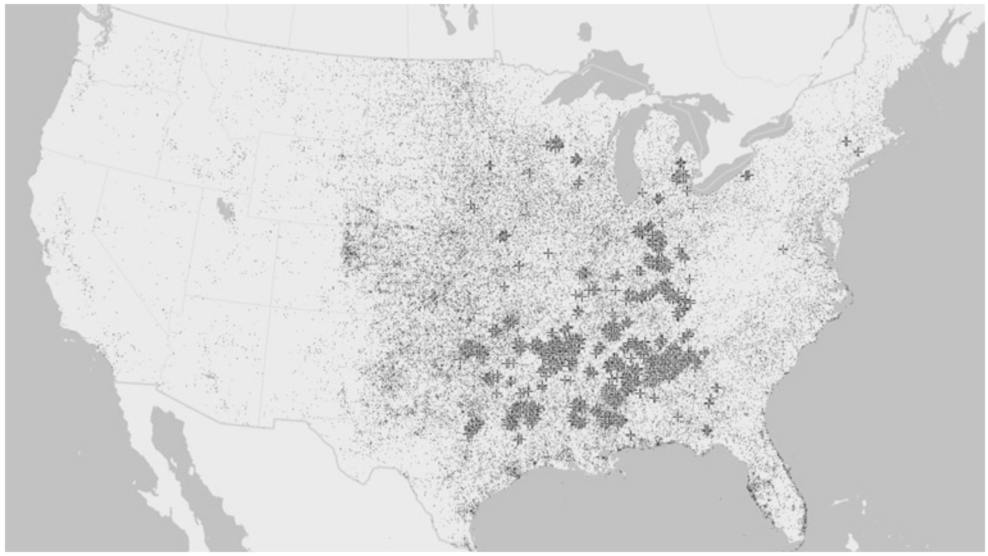


Fig. 5.3 Tornado hotspots (+) and reported start points across the United States from 1950–2018; hotspots are calculated using the Getis-Ord Gi* statistic on tornado geographic frequency and weighted by severity (Fugita Scale 0–5) to determine locations with a higher risk of damage based upon reported historical events (Pvalue < 0.05 & Zscore > 3). Tornado data from the NOAA Storm Prediction Center for Severe Weather

- *Similarity* – identify the features that are either the most similar or least similar to another set of features based upon attribution.

5.4.3 Pattern Analysis

Pattern analysis involves identifying, quantifying, and visualizing spatial patterns in spatial data [17, 18]. Identifying geographic patterns is important for understanding how geographic phenomena behave.

Although it is possible to understand the overall pattern of features and their associated values through traditional mapping, calculating a statistic quantifies the pattern [19]. Statistical quantification facilitates the comparison of patterns with different distributions or across different time periods. Pattern analysis tools are often used as a starting point for more in-depth analyses. For example, spatial autocorrelation can be used to identify distances where the processes promoting spatial clustering are most pronounced. This might help the user to select an appropriate distance (scale of analysis) to use for investigating hot spots (hot spot analysis using the Getis-Ord Gi* statistic, see Fig. 5.3).

Pattern analysis tools are used for inferential statistics; they start with the null hypothesis that features, or the values associated with the features, exhibit a spatially random pattern. They then compute a p-value representing the probability that the null hypothesis is correct (that the observed pattern is simply one of many possible versions of complete spatial randomness). Calculating a probability may be important if you need to have a high level of confidence in decision making. If there are public safety or legal implications associated with your decision, for example, you may need to justify your decision using statistical evidence.

5.4.4 Cluster Analysis

Cluster analysis is used to identify the locations of statistically significant hot spots, spatial outliers, and similar features [20]. Cluster analysis is particularly useful when action is needed based on the location of one or more clusters. An example would be the assignment of additional police officers to deal with a cluster of burglaries. Pinpointing the location of spatial clusters is also important when looking for potential causes of clustering; where a disease outbreak occurs can often provide clues about what might be causing it. Unlike pattern analysis (which as used answer the questions such as "Is there spatial clustering?") cluster analysis supports the visualization of the cluster locations and extent. Cluster analysis can be used to answer the questions such as "Where are the clusters (hot spots and cold spots)?", "Where are incidents most dense?", "Where are the spatial outliers?", and "Which features are most alike?" (Fig. 5.4).

5.4.5 Proximity Analysis

Proximity analysis allow people to answer one of the most common questions posed in spatial analysis: "What is near what?". This type of analysis supports the determination of proximal of features within one or more datasets; e. g., identify features that are closest to one another or calculate the distances between or around them. Common analysis methods include

- *Distance calculation* – the Euclidean distance from a single source or set of sources.
- *Travel cost calculation* – the least accumulative cost distance from or to the least-cost source, while accounting

Fig. 5.4 Spatio-temporal clustering (DBSCAN) of ridesharing drop off locations in Midtown Manhattan. This identified clusters (darker points in the figure) where many drop offs occurred in a similar place and time and the minimal cluster size is 15 events

for surface distance along with horizontal and vertical cost factors.

- *Optimal travel cost calculation* – the optimum cost network from a set of input regions. One example application of this tool is finding the best network for emergency vehicles.

5.4.6 Predictive Modeling

Predictive analytics build models to forecast behavior and other future developments. It encompasses techniques from spatial statistics, data mining, machine learning, and artificial intelligence [21–23]. Patterns are identified in historical data and are used when creating models for future events.

Machine learning uses algorithms and statistical models to analyze large data sets without using explicit sequences of instructions. Machine learning algorithms create a model of training data that is used to make optimized predictions and decisions. Machine learning is considered to be a subset of artificial intelligence.

Deep learning is a subset of artificial intelligence where models resembling biological nervous systems are arrayed in multiple layers where each layer uses the output of the preceding as input to create a more abstract and composite representation of the data [24]. Deep learning architectures include deep neural networks, belief networks and recurrent neural networks. Deep learning is commonly used in the domains of natural language processing, computer vision, and speech recognition.

5.5 Technology and Tools

There are several key technologies that are commonly employed to process large volumes of spatial data. These technologies are frequently distributed in nature, allowing collections of computing resources to work collaboratively toward solutions. This collection includes distributed processing frameworks and distributed data stores. At the most basic level, distributed systems are collections of networked computers that work in a coordinated manner; this is sometimes termed "concurrent computing", "parallel computing", or "distributed computing". In a distributed computing environment, the processors run concurrently and each processor has its own private memory (distributed memory). Information is exchanged between processors through messages between the processors. In parallel computing, all processors in the cluster usually have access to shared memory that is used to exchange information between processors.

A distributed data store is a computer cluster where data is persisted on more than one node, often in a replicated fashion. Distributed databases are commonly non-relational databases that are optimized to support rapid or parallel access of data across a large number of nodes. Distributed databases usually expose rich query capabilities; some however are limited to a key-value store semantics. Examples of distributed databases are Google's Bigtable [25], Amazon's Dynamo, and Windows Azure Storage.

5.5.1 Available Tools

There are a number of tools and technologies that are used to support big data analytics. Some of the tools are open source, while others are more traditional commercial offerings. This collection encompasses distributed file systems, distributed processing frameworks, NoSQL and columnar data stores, as well as cloud-based computational platforms [26].

Distributed Processing Frameworks

Apache Hadoop is an open-source software framework that allows a cluster of commodity computers to solve problems involving large amounts of data and/or computation. Hadoop was motivated by work at Google on the MapReduce programming model and the Google File System (GFS) [27]. Hadoop supports a distributed storage and processing framework using the MapReduce programming model [28]. Spark is a more modern open-source, distributed processing framework that is optimized to support running large-scale data analytics applications across clustered system; it differs from MapReduce as it better supports in-memory and memory pipelined applications.

Hadoop was designed for computer clusters built from commodity hardware; this follows the original Google GFS model. Hadoop is designed with the assumption that hardware failures are common and should be automatically handled by the framework. The core of Hadoop is the Hadoop Distributed File System (HDFS) [29], a resource manager, and a distributed processing framework that supports the MapReduce programming model. HDFS splits large files into shards (or blocks) that are distributed across multiple nodes in a cluster. Reliability is achieved through data replication; copying or replicating the blocks across multiple nodes in the cluster. Hadoop distributes software across the collection of nodes in the cluster to enable the processing of data in parallel. Take the compute to the data allows large datasets to be processed faster and more efficiently than in systems that rely on a parallel file systems where the compute and the data are distributed via high-speed communication architectures.

Hadoop can be deployed in on-premise datacenters as well as in the cloud; this allows organizations to deploy Hadoop without acquiring expensive hardware or having setup and operational (dev-op) expertise. Hadoop compatible cloud offerings are available from Amazon, Microsoft, IBM, Google, and Oracle (among others).

Datastores

NoSQL databases (originally referencing “non-SQL” or “non-relational”) store and retrieve data differently from standard relational databases. NoSQL databases (sometimes considered next generation databases) are designed to address some of the limitations of traditional relational databases such as being distributable, simpler in design, often open-source, and horizontally scalable. Many databases supporting these characteristics originated in the late 1960s; the “NoSQL” description was employed beginning in the late 1990s with the requirements imposed by companies such as Facebook, Google, and Amazon. NoSQL databases are commonly used with big data applications. NoSQL systems are also sometimes called “Not only SQL” to emphasize that they may support SQL-like query languages.

In order to achieve increased performance and scalability, NoSQL databases commonly used data structures (e. g., key-value, columnar, document, or graph) that are different from those used in relational databases. NoSQL databases vary in terms of applicability to particular problem domains. NoSQL databases are often classified by their primal data structures; examples include:

- *Key-value:* Apache Ignite, Couchbase, Dynamo, Oracle NoSQL Database, Redis, Riak
- *Columnar:* Accumulo, Cassandra, Druid, HBase, Vertica
- *Document:* Apache CouchDB, Cosmos DB, IBM Domino, MarkLogic, MongoDB
- *Graph:* AllegroGraph, Apache Giraph, MarkLogic, Neo4j, Spark GraphX
- *Multi-model:* Apache Ignite, Couchbase, MarkLogic

Cloud Platforms

Big data analytic systems were often deployed on premises; however, cloud platform vendors have made it easier deploy big data systems in the cloud. Cloud-based services enable organization to create cloud-based clusters and run analytical processes as long as necessary. These clusters can then be taken offline when they are no longer needed [30]. Cloud platforms commonly support scaling horizontally (also termed scale-out, e. g., adding more nodes to a system), as well as vertically (scale-up, e. g., adding resources to a node, such as CPU cores, memory, or storage).

Platform as a Service (PaaS) is a category of cloud computing services that provides a platform allowing organizations to run and manage distributed applications without the complexity of building and maintaining the infrastructure usually associated with developing and launching an application. PaaS is commonly delivered in one of three ways:

- as a public cloud service from a provider,
- as a private service (on-premises) inside the firewall, or
- as software deployed on a public infrastructure as a service.

Big Data as a Service (BDaaS) is a new concept that combines Software as a Service (SaaS), Platform as a Service (PaaS), and Data as a Service (DaaS) to address the requirements of working with massively large data sets. BDaaS offerings commonly incorporate the Hadoop stack (e. g., HDFS, Hive, MapReduce, Pig, Storm, and Spark), NoSQL data stores, and stream processing capabilities.

Microsoft Azure is a cloud computing service utilizing Microsoft-managed data centers that supports both software as a service (SaaS) and platform as a service (PaaS); visu-

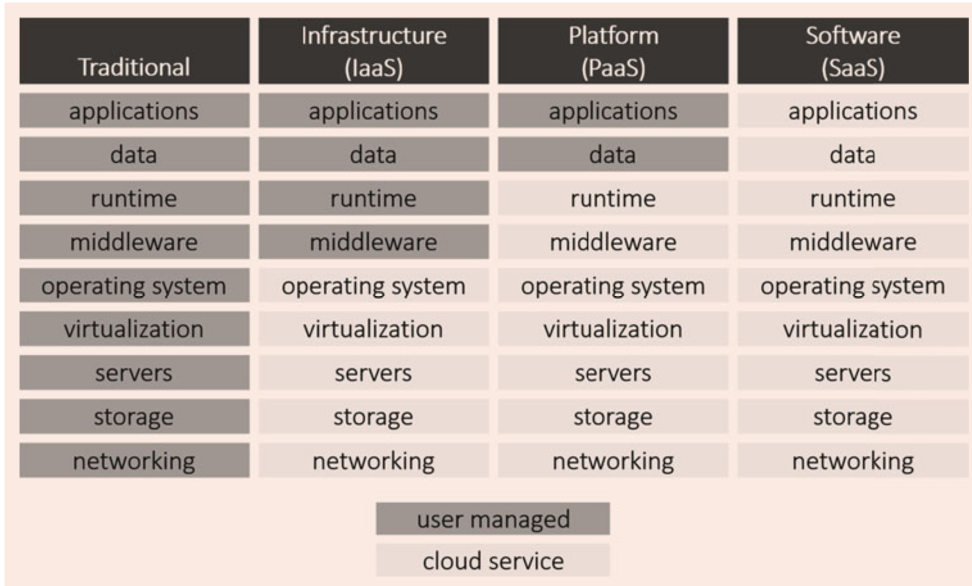


Fig. 5.5 Cloud service models (IaaS, PaaS, and SaaS) [31]

alization of the differences between the main cloud service models is provided in Fig. 5.5. It provides data storage capabilities including Cosmos DB (a NoSQL database), the Azure Data Lake, and SQL Server-based databases. Azure supports a scalable event processing engine and a machine learning service that supports predictive analytics and data science applications.

The Google Cloud is a PaaS offering that supports big data with data warehousing, batch and stream processing, data exploration, and support for the Hadoop/Spark framework. Key components include BigQuery, a managed data warehouse supporting analytics at scale, Cloud Dataflow, which supports both stream and batch processing, and Cloud Dataproc, a framework for running Apache MapReduce and Spark processes.

Amazon AWS, though commonly considered an Infrastructure as a Service (IaaS) where the user is responsible for configuration, also provides PaaS functionality. Amazon supports Elastic MapReduce (EMR) that works in conjunction with EC2 (Elastic Compute Cloud) and S3 (Simple Storage Service). Data storage is provided through DynamoDB (NoSQL), Redshift (columnar), and RDS (relational data store). Machine learning and real-time data processing infrastructures are also supported.

Other significant examples of BDaaS providers include the IBM Cloud and the Oracle Data Cloud. Big data Infrastructure as a Service (IaaS) offerings (that work with other clouds such as AWS, Azure, and Oracle) are also available from Cloudera, and Databricks.

GIS – Hadoop-GIS, SpatialHadoop, Esri GeoAnalytics Server

In the academic and commercial realms, there are several systems of note. Hadoop-GIS [32] is an academic distributed spatial data warehousing and query processing system that utilizes the Hadoop along with the MapReduce programming model. Hadoop-GIS supports spatial partitioning and exposes a customizable spatial query engine (RESQUE) along with the ability to perform 2-D and 3-D spatial joins. Declarative queries are supported via an integration with Hive. A successor was implemented on Spark, called SparkGIS. SparkGIS also supports spatially aware management of partitions loaded into memory rather than arbitrary spilling to disk. It was benchmarked using medical pathology images as well as OpenStreetMap data.

SpatialHadoop is another academic research system [33]; it is an open source MapReduce extension to Hadoop that is focused on big spatial data. It has a custom spatial high-level language along with support for native spatial data types, spatial indexes (grid files, R-trees, and R+-trees), and spatial query operations (e. g., range queries, kNN, and spatial joins) on HDFS.

Other interesting academic research systems include GeoSpark, a framework for performing spatial joins, range queries, and kNN queries. GeoSpark supports both quadtree and R-tree based indexing of spatial data [34]. GeoSpark uses a regular grid for global partitioning, with local spatial indexes. Simba is another system that provides range, distance, and kNN queries and joins. It uses two-level indexing and can support custom partitioning of data [35]. Simba does not support spatio-temporal queries. Magellan is open-source software for spatial analytics based upon Spark [36]. It supports Spark SQL for traditional SQL processing as well as a custom broadcast join. It uses the Java API provided by GIS Tools for Hadoop [37]. LocationSpark is another Spark-based library that supports range queries, spatial joins, and kNN queries [38]. Spatial data is stored in key-value pairs with a geometry key. GeoMesa is a framework built upon

Accumulo that provides geohash-based spatial indexing and query capabilities [39]. Finally, STARK is another Spark-based framework that supports range queries, kNN queries, and range queries on both spatial and spatio-temporal data. STARK also support density-based spatial clustering (DBSCAN) [20, 40].

The Esri GeoAnalytics Server is a big spatial and temporal data processing and analysis capability of the ArcGIS Enterprise platform. It utilizes the Spark distributed processing framework to support aggregation, regression, clustering, and analysis of big spatial data [41]. It works with distributed file shares, HDFS, cloud storage, and Hive. It provides a large collection of tools that can be accessed through the ArcGIS desktop, the enterprise portal Map Viewer, a REST API, or via Python directly.

5.6 Challenges

The huge volumes of spatial data coupled with its variety cause significant data management challenges with data quality, consistency and governance [42]. Building and maintaining the diverse collection of commercial and open-source big data processing tools and architectures (e. g., Apache Hadoop, HDFS, and Spark) in an accessible and cohesive architecture is a challenging proposition for most organizations. Another common problem when organizations initiate big data analytics initiatives include a lack of analytics skills among existing personnel coupled with the high cost of hiring new data scientists.

Recently, the proliferation and advancement of AI and machine learning technologies have enabled vendors to produce software for big data analysis that is easier to use, particularly for the growing citizen data scientist population. Some of the leading vendors in this field include Alteryx, IBM, and Microsoft.

The major challenges of spatial big data and analytics is less about the hardware and more about identifying individuals that are capable of working with and managing large volumes of data and being able analyze it and identify information that is valuable to their organizations.

The complexities related to the relationship between hardware, software, and expertise has evolved with time. The cost of hardware (CPUs and storage) was an original big data challenge. During the past decade, the cost per gigabyte for computer storage has dropped by a factor of five. Similar trends in processing power, memory, and communication infrastructures have also been observed.

Most organizations can afford big data processing hardware that will support the storage and analytic processing; smaller organizations can alternatively employ highly scalable cloud solutions that will support their spatial big data analytic requirements.

5.7 Summary

Big data analytics on spatial data (e. g., moving sensors, aerial and satellite imagery, Lidar, social networks, etc.) commonly involves spatial processing, sophisticated spatial statistical algorithms, and predictive modeling. GIS users and data scientists apply big data analytics to evaluate these large collections of data; data with volumes that traditional analytical systems are unable to accommodate. Spatial big data is differentiated from standard big data by the presence of spatial relationships, geostatistical correlations, and spatial semantic relations; these additional challenges are beyond what is usually encountered with traditional big data. When working with big data, common analytic objectives and workflows include the visualization and identification of patterns and trends, filtering and converting streaming data contains geographical elements into geographic layers of information, cluster and proximity analysis, and predictive modeling.

To address these demands, new analytic environments and technologies have been developed; this includes distributed processing infrastructures (e. g., Hadoop and Spark), distributed file stores (e. g., HDFS), NoSQL databases (e. g., Accumulo, Cassandra, Dynamo, HBase, and MongoDB), and big data enabled cloud platforms (e. g., Azure, Amazon, Google, SAP, and Oracle).

The combination of recent developments in advanced analytical processing and sophisticated distributed processing technologies and infrastructures enables both modest and large organizations to take advantage of spatial big data and obtain new insights and understanding of their problem domains and communities.

References

1. Miller, H., Goodchild, M.: Data-driven geography. GeoJournal **80**(4), 449–461 (2014). https://doi.org/10.1007/s10708-014-9602-6
2. Dean, J., Ghemawat, S.: Mapreduce: simplified data processing on large clusters. Commun. ACM **51**(1), 107–113 (2008). https://doi.org/10.1145/1327452.1327492
3. Garillot, F., Maas, G.: Stream Processing with Apache Spark: Best Practices for Scaling and Optimizing Apache Spark. O'Reilly Media, Sebastopol (2018)
4. Zaharia, M., Chowdhury, M., Franklin, M., Shenker, S., Stoica, I.: Spark: cluster computing with working sets. In: Proc. of the 2nd USENIX conference on Hot Topics in Cloud Computing (HotCloud 2010) (2010)
5. Alexander, W., Copeland, G.: Process and dataflow control in distributed data-intensive systems. In: Proc. of the 1988 ACM SIGMOD International Conference on Management of Data (SIGMOD '88), pp. 90–98. (1988) https://doi.org/10.1145/50202.50212
6. DeWitt, D., Gray, J.: Parallel database systems: the future of high performance database systems. Commun. ACM (1992). https://doi.org/10.1145/129888.129894
7. Klein, J., Buglak, R., Blockow, D., Wuttke, T., Cooper, B.: A reference architecture for big data systems in the national security

domain. In: Proc. of the 2nd International Workshop on BIG Data Software Engineering (BIGDSE 2016) (2016) https://doi.org/10.1145/2896825.2896834
8. NoSQL: NoSQL NoSQL definition. https://www.nosql-database.org
9. Pavlo, A., Aslett, M.: What's really new with NewSQL? SIGMOD Rec. **45**(2), 45–55 (2016). https://doi.org/10.1145/3003665.3003674
10. Apache: Welcome to Apache Hadoop! (2018). https://hadoop.apache.org
11. Marz, N., Warren, J.: Big Data: Principles and Best Practices of Scalable Realtime Data Systems. Manning, Greenwich (2013)
12. Mysore, D., Khupat, S., Jain, S.: Big data architecture and patterns (IBM White Paper) (2013). https://www.ibm.com/developerworks/library/bdarchpatterns1
13. Hägerstrand, T.: What about people in regional science? Pap. Reg. Sci. Assoc. **24**(1), 7–24 (1970). https://doi.org/10.1111/j.1435-5597.1970.tb01464.x
14. Barwick, H.: The 'Four Vs' of Big Data, Implementing Information Infrastructure Symp., Computerworld. https://www.computerworld.com.au/article/396198/iiis_four_vs_big_data
15. Hey, A., Tansley, S., Tolle, K.: The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research, (2009)
16. Longley, P., Goodchild, M., Maguire, D., Rhind, D.: Geographic Information Systems and Science, 2nd edn. Wiley, Hoboken (2015)
17. Bonham-Carter, G.: Geographic Information Systems for Geoscientists: Modelling with GIS. Pergamon, New York (1994)
18. Golledge, R., Stimson, R.: Spatial Behavior: A Geographic Perspective. Guilford, New York (1997)
19. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, Berlin, Heidelberg, New York (2000) https://doi.org/10.1007/978-1-4757-3264-1
20. Ester, M., Kriegel, H., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD 1996), pp. 226–231. (1996)
21. Minsky, M.: The Society of Mind. Simon & Schuster, New York (1986)
22. Newell, A., Shaw, J., Simon, H.: Report on a general problem-solving program. Commun. ACM **2**(7), 256–264 (1959)
23. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Perrot, M., Duchesnay, É.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. 2825–2830 (2011)
24. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature14539
25. Chang, F., Dean, J., Ghemawat, S., Hsieh, W., Wallach, D., Burrows, M., Chandra, T., Fikes, A., Gruber, R.: Bigtable: a distributed storage system for structured data. Acm Trans. Comput. Syst. **26**(2), 1–26 (2008). https://doi.org/10.1145/1365815.1365816
26. Sena, B., Allian, A., Nakagawa, E.: Characterizing Big Data Software Architectures: A Systematic Mapping Study. In: Proc. of the 11th Brazilian Symposium on Software Components, Architectures, and Reuse (SBCARS 2017) (2017) https://doi.org/10.1145/3132498.3132510
27. Ghemawat, S., Gobioff, H., Leung, S.: The Google file system. In: Proc. of the 19th ACM Symposium on Operating Systems Principles, pp. 29–43. (2003) https://doi.org/10.1145/945445.945450
28. Sakr, S., Liu, A., Fayoumi, A.: The family of mapreduce and large-scale data processing systems. ACM Comput. Surv. **46**(1), 1–44 (2013). https://doi.org/10.1145/2522968.2522979
29. Shvachko, K., Kuang, H., Radia, S., Chansler, R.: The Hadoop distributed file system. In: Proc. of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST 2010) (2010) https://doi.org/10.1109/MSST.2010.5496972
30. Chang, W., Abu-Amara, H., Sanford, J.: Transforming Enterprise Cloud Services. Springer, London (2010)
31. Chou, D.: Cloud service models (IaaS, PaaS, SaaS) diagram. https://dachou.github.io/2018/09/28/cloud-service-models.html
32. Aji, A., Wang, F., Vo, H., Lee, R., Liu, Q., Zhang, X., Saltz, J.: Hadoop-GIS: a high performance spatial data warehousing system over mapreduce. Proc. VLDB Endow. **6**(11), 1009–1020 (2013)
33. Eldawy, A., Mokbel, M.: SpatialHadoop: a mapreduce framework for spatial data. In: Proc. IEEE 31st International Conference on Data Engineering (ICDE 2015), pp. 1352–1363. (2015)
34. Yu, J., Wu, J., Sarwat, M.: Geospark: a cluster computing framework for processing large-scale spatial data. In: Proc. of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL 2013) (2013) https://doi.org/10.1145/2820783.2820860
35. Xie, D., Li, F., Yao, B., Li, G., Zhou, L., Guo, M.: Simba: efficient in-memory spatial analytics. In: Proc. of the 2016 International Conference on Management of Data (SIGMOD 2016), pp. 1071–1085. (2016) https://doi.org/10.1145/2882903.2915237
36. Sriharsha, R.: Magellan: geospatial analytics on spark (2017). https://www.hortonworks.com/blog/magellan-geospatial-analytics-in-spark
37. Whitman, R., Park, M., Ambrose, S., Hoel, E.: Spatial indexing and analytics on Hadoop. In: Proc. of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL 2014), pp. 73–82. (2014) https://doi.org/10.1145/2666310.2666387
38. Tang, M., Yu, Y., Malluhi, Q., Ouzzani, M., Aref, W.: LocationSpark: a distributed in-memory data management system for big spatial data. Proc. VLDB Endow. **9**(13), 1565–1568 (2016). https://doi.org/10.14778/3007263.3007310
39. Hughes, J., Annex, A., Eichelberger, C., Fox, A., Hulbert, A., Ronquest, M.: Geomesa: a distributed architecture for spatio-temporal fusion. In: Proc. of SPIE Defense and Security (2015) https://doi.org/10.1117/12.2177233
40. Hagedorn, S., Götze, P., Sattler, K.: The STARK framework for spatio-temporal data analytics on spark. In: Proc. of the 17th Conference on Database Systems for Business, Technology, and the Web (BTW 2017) Stuttgart, Germany (2017)
41. Whitman, R., Park, M., Marsh, B., Hoel, E.: Distributed spatial and spatio-temporal join on Apache spark. Acm Trans. Spatial Algorithms Syst. **5**(1), 1–28 (2019). https://doi.org/10.1145/3325135
42. Hilbert, M.: Big data for development: a review of promises and challenges. Dev Policy Rev **34**(1), 135–174 (2015)
43. Abel, D., Ooi, B., Tan, K., Power, R., Yu, J.: Spatial join strategies in distributed spatial DBMS. In: Proc. Advances in Spatial Databases – 4th International Symposium, SSD'95 (Portland, ME 1995) Lecture notes in computer science, vol. 1619, pp. 348–367. Springer, Berlin, Heidelberg, New York (1995)

Erik Hoel Erik Hoel received his PhD from the University of Maryland at College Park in 1995. He works at Esri in Redlands, CA. His focus is geodata management, parallel and distributed processing, big data, network models (transportation, utility, and telco), and large service-based architectures.

Part B

Geographic Information

Geographic information may be subdivided in many ways. This handbook follows a classic view and sets geodesy, surveying, cartography, photogrammetry, and remote sensing as the prime chapters supplemented by an introduction (Geographic Information Systems), and technical topics, namely standardization, geometry/topology, metadata, geospatial web services, geosemantic interoperability, registration, security, and change detection.

Norbert Bartelme has compiled the introduction. Several authors worked on the prime chapters. Mathias Becker provides a compact overview of geodesy. Vladimir Golubev introduces methods and equipment in surveying. Aileen Buckley, Paul Hardy, and Kenneth Field provide a comprehensive and detailed paper on modern cartography, Jan Skaloud with Michael Cramer and Norbert Haala explain the mathematics of photogrammetry based on the modern approach from computer vision, and Erik Borg, Sina Truckenbrodt, Angela Lausch, Peter Dietrich, as well as Karsten Schmidt provide a comprehensive coverage of modern remote sensing.

Standardization takes into account the ISO/TC 211 standards (Wolfgang Kresse) and the Open Geospatial Consortium (OGC) work (David Danko). The chapter on geometry and topology follows a generic concept and is written by Gerhard Gröger and Betsy George respectively. The metadata chapter's author is David Danko who has guided the development of the ISO metadata standard over two decades. Joan Masó being involved in the OGC development for more than a decade explains OGC's geospatial web services in his chapter. Jean Brodeur being a forerunner of the ISO-standard for geosemantic interoperability is author of the chapter that discusses the role of formalized geography-terms across language barriers. Douglas O'Brien and Roger Lott explain registries for geospatial parameters as essential tools for the homogeneous worldwide processing of spatial data. A common topic many of us worry about is the security of personal data; Andreas Matheus covers this topic including spatial data. Automatic change detection for the update of maps has always been a topic of interest in remote sensing and map production; Jérôme Théau is the author of this chapter.

6 Geographic Information Systems

Norbert Bartelme

Contents

Abstract

Before addressing the diverse functionalities that can be found in geographic information systems (GISs) today, this chapter defines what a geographic information system is, its purpose, and its general architecture. In Sect. 6.1, the different forms of GISs that are found in the ever-expanding range of information technology tools are discussed. Dedicated GIS systems in the strictest sense are more and more being complemented by geographically aware applications embedded in diverse environments such as mobile. Interoperability in structure and semantics has become a central issue, bridging the gaps between traditional GIS services, social media, user-driven information supply, and the upcoming Internet of Things. The term *GIS* therefore serves as a conceptual headline for an ever-increasing scope of tools. The core of each GIS is constituted by analysis functions. They are the reason why a GIS is created in the first place. The second part of this chapter (Sect. 6.2) contains a list of the most common GIS functionality categories. A few typical examples of such analysis functions are described in more detail to explain what geographic information (GI) is and how its digital form can be utilized to solve problems of a geospatial nature efficiently, to gain insight into the processes of a geospatial nature that influence many aspects of our life, and to arrive at decisions that are sound, explainable, and repeatable.

Keywords

GIS · geographic information · geospatial · modeling · database · interoperability · Internet of Things · Linked Data · analysis functions

6.1 Architecture of a GIS

6.1.1 Information and Data

Before it is possible to provide a definition for a *geographic information system* (GIS) or, using a term that has become increasingly popular, a *geospatial information system*, we must first clarify what *information* means in the sense of information technology (IT), and how and to what degree it specializes into geographic/geospatial information. The basics for this discussion can be found in the introductory chapter of this book, where modeling and encoding are dealt with in detail and within the broad spectrum of information technology, meaning that it is also relevant for geographic information technology (Chaps. 1 and 4). The reader may wonder why we place such importance on the definition of the term *information*, since it is used widely in connection with—and sometimes identically to—digital data, which are becoming ubiquitous in our world. The Internet, for example, can be seen as an ever-expanding resource of digital data,

N. Bartelme (✉)
Institute of Geodesy, Graz University of Technology
Graz, Austria
e-mail: norbert.bartelme@tugraz.at

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_6

Fig. 6.1 Data (points, lines, areas) and information (streets, rivers, etc.). (Source for base data: City of Graz Geoportal. Overlay: Institute of Geodesy, TU Graz)

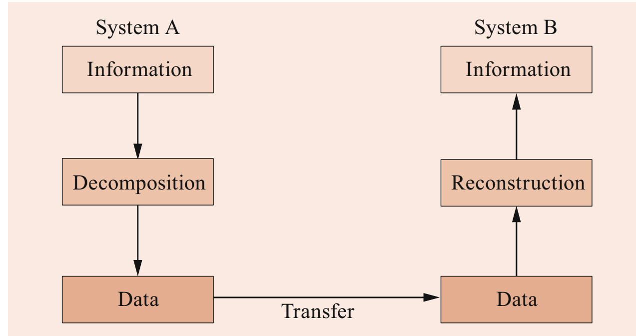


Fig. 6.2 Transfer of data and (implicitly) of information (Fig. 4.1 in Chap. 4)

much of which is geographic data. From this pool of data, we seek to retrieve information that will answer the questions or satisfy the criteria that made us search. The crucial difference between data and information is already visible. Any source on the Internet, any database, can only contain digital data. Information is generated in our brains, when we interpret the data to solve our questions or problems. The Latin root of the word is *informare*, which literally means *to give form or shape,* and this is usually extended into *to give form to the mind,* as in education, instruction, or training. In contrast to data, information can be seen as an answer to a question (even if it has only been implicitly posed) that heightens the level of understanding of the inquirers or makes them capable of reaching a goal.

Figure 6.1 illustrates data—pixels or points, lines, and areas—from which we retrieve the information that these data describe part of a city, with streets, a river, and other topographic landmarks. If we utilize a web tool for route planning, wanting to go from street A to square B, such services usually answer with maps and textual descriptions of the recommended route. Example data are the characters in a string of such text, or paragraphs, or the whole text. Data can also be individual pixels, or chunks of pixels, or a rectangular array of pixels constituting the map. Information is what we retrieve from the text, including its meaning for us: the recommended roads, turns, and routes. Information is also what we see on the map image. Our eyes register pixels (data), but our mind registers the best route (information). A third, still higher, level on top of data and information would be knowledge: a concept that extends the concept of information into comparing, remembering, and learning. If one recognizes the city shown in Fig. 6.1, one has crossed from information to knowledge. Also, in the routing example just given, a knowledge engineer would combine the collective memory of typical route users, such as taxi drivers, and use their experience (which is sometimes difficult to model) about typical recommendations for times of day, seasons, or vehicle types to arrive at a complex knowledge system that becomes smarter with every use.

Returning to the difference between and the blending of data and information, we arrive at a point that is crucial for information technology, for any information system, and therefore also for geographic/geospatial information technology implemented in a GIS or—in a web-enabled environment—in *GI services*. One of the things that information systems have to do is to transport information from system A to system B or from user C to user D; yet, really, only data can be transported (Fig. 6.2). This means that information in the sending system has to be decomposed into chunks of data that can be—and must be—transported to the receiving system, where they can be reconstructed into information. The last, and most relevant, reconstruction step has to be taken by the user sitting in front of the screen in reading and interpreting text or a map. It is clear that this process, which proceeds in opposite directions at the sending and receiving ends, can only result in equivalent information content on both sides in theory. In all practical cases, information loss has to be taken into account. The more elaborate the information construct, the greater the danger of information loss. Information that is highly interconnected in a network, information that is highly structured, and information carrying strong application-dependent semantics will be especially vulnerable in such a sequential stream of basic data entities. One of the most relevant consequences for the design of any information system is therefore the necessity to place great importance on the ways in which users are assisted when performing this reconstruction of data into information they need. Of course, data may also show a structure; for example, a sequence of value pairs carrying some common attributes. Reading the attribute values *River* and *Danube* and noting that the value pairs are given by real numbers, we interpret this data as the River Danube.

Likewise, a linear concatenation of blue pixels on the screen may spawn the same interpretation, leading us from data to information.

To sum up, the difference between information and data relevant to a GIS is characterized as:

- *Data* is what is stored and transported, such as strings of characters or pixels, or defined structures thereof
- *Information* is a result of the interpretation performed when visualizing or analyzing data.

6.1.2 Geographic (Geospatial) Information

Having clarified the concepts of and the differences between data and information, we can proceed to geographic information, or geoinformation, or geospatial information. The term *geoinformation* is mostly used in German-speaking countries. *Geospatial* is a term that successfully tries to bridge the gap between geographic information in the strictest sense (for example, digital terrains) and spatial information (for example, a model for a three-dimensional (3-D) structure, a building, or a bridge). In a widened perspective, this term denotes not only things that exist (or are being planned) at some location on the Earth's surface, but also events such as traffic congestions, floods, and, yes, also events in everyday language, such as an open-air festival. For the rest of this chapter and for the sake of simplicity, let us assume that the terms mentioned are all synonymous. All these examples share one basic aspect: they exist or happen somewhere on the Earth's surface and they have a spatial extent; in many cases they also have a temporal aspect, a position in time, and a temporal extent. We can therefore speak of a four-dimensional (4-D) continuum (three spatial dimensions and one temporal dimension) that characterizes geographic data and geographic/geospatial information.

At this point in the discussion, we may stop and consider what types of information are nonspatial or nongeographic in the sense that they are independent of where and when they are relevant. Strictly speaking, there is hardly any such information that is totally void of any geospatial context. This context may not always appear in the model of such information, but it is inherently there. In cases of land use, traffic and transport, climate, agriculture, and economics, the spatial aspect is evident and, in many cases, also modeled. In other cases, such as linguistics, philosophy, and literature, it may be less evident and is seldom modeled. At the end of the range are laws of mathematics, physics, and chemistry, which are almost independent of the location where their validity is tested. This seemingly philosophical discussion about the extent of geospatially relevant application domains—highbrow as it seems at first sight—has some very practical consequences for the design of a GIS. The more straightforward the geospatial interpretation, the clearer the modeling of the geometric and topological properties of such information. Well-known geometric entities exist for land use and cadastre applications, and this is also the case for roads and intersections in traffic and transport. However, the geospatial extent of a dialect in linguistic GIS applications requires a different approach, since this extent is fuzzy.

To sum up, geographic (or geospatial) information is characterized as:

- *Location, extent, and coverage* are aspects of prime importance for such information
- However, *geometrical concepts* may—depending on the application—be concise or fuzzy.

6

6.1.3 Geographic Information System Definitions

In its early days, geospatial information was handled by a small number of dedicated systems called GISs, serving a well-defined community of expert users. Such systems can still be found in large public administrative bodies, governmental agencies, and utility/transport network authorities. However, in the past two decades, a big shift towards embedding some GIS functionality as GI services in mainstream IT has become noticeable. The advent of ubiquitous and fast web access and of ever more powerful mobile devices has opened up a mass market. Today, any smartphone can, to some extent, be seen as a GIS. Likewise, the range of potential users has grown from an originally small expert group to most anyone today. Therefore, a definition for the essence of a GIS can only be given on a general conceptual level. Most of the following definitions were formulated in the early GIS days, but they have survived the transition phases and are still as valid as they used to be. The choice of definitions used here reflects the fact that each of the following citations carries a flavor that distinguishes it to some extent from the other definitions.

- A GIS captures, stores, analyzes, and visualizes data that describe a part of the Earth's surface, technical and administrative entities, as well as findings from geoscience, economics, and ecological applications [1, 2]
- It is an information system with a database of observables of spatially distributed objects, activities, or events that can be described by points, lines, or surfaces [3]
- It is a comprehensive collection of tools for the capture, storage, retrieval, transformation, and visualization of spatial data on the real world for special applications [4]
- It is an information system containing all spatial data on the atmosphere, the Earth's surface, and the lithosphere,

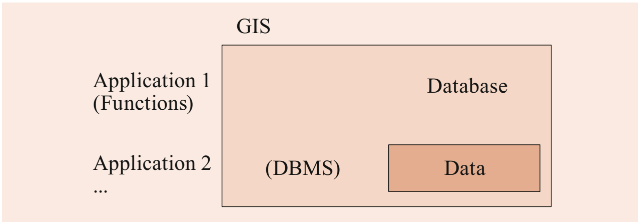


Fig. 6.3 A GIS in the classical sense

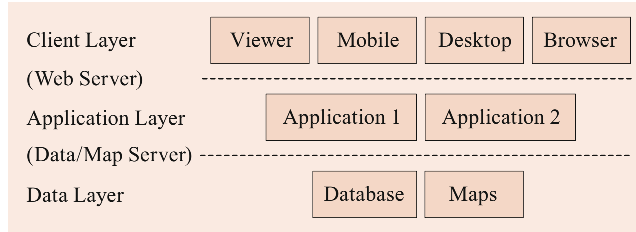


Fig. 6.4 A GIS in a client–server architecture

allowing the systematic capture, updating, manipulation, and analysis of such data based on a standardized reference frame [5]
- It is a system for decision support that integrates spatial data in a problem-solving environment [6].

Further definitions of a GIS may be found in [7]. Depending on the point of view, a GIS can be seen as:

- A collection of geospatial data plus the corresponding functions for storage and retrieval
- A collection of algorithmic and functional tools
- A set of hardware and software components needed for the handling of geospatial data
- A special type of information technology
- A gold mine for answers to geospatial questions
- A model of spatial relationships and spatial reconnaissance.

6.1.4 Classical GISs and Recent Modifications

In the classical sense, an information system consists of a database representing the inner core of the system, which is managed by a database management system (DBMS), and an outer shell of tools that can be utilized by the user for manipulating and analyzing this data (Fig. 6.3). This definition, taken from [8], is one of the earliest and best known. Chapter 3 is devoted to databases; the reader is invited to check further details there. Here, we concentrate on the impact of databases on GISs. Using the definition just given, traditional GISs conform to this concept. In the early years of GIS technology, the database usually followed a standard relational database approach, but this concept has subsequently been expanded into an *object-relational database management system* (ORDBMS) to enable geospatial data types and geospatial predicates and to provide the means to deal with the ever-growing volume of geospatial data collections by supplying methods for organizing, searching, and retrieving large datasets.

There is no such thing as a general-purpose information system. Data are abstractions of reality. An abstraction process can lead to different results, depending on the view of the abstractor. Therefore, each information system is *application dependent*. The degree of dependence varies for the different components of an information system. The data will often lend themselves to more than one purpose. This is also advisable for reasons of practicality, efficiency, and cost. The functions will depend more heavily on the application that is envisioned. Typically, a GIS offers functions for storage and retrieval, performing queries, visualization, transformation, geometrical and thematic analysis, and more (see also the next section for more details). Even though these functions are more or less available in every GIS, their specific forms depend to a large extent on the application. As an example, visualization in a cadastral application will have prerequisites that are quite different from those for a 3-D city model. The same holds for the other types of functions.

The traditional setup of a GIS has been modified in several ways due to the arrival of new technologies and new concepts. The arrival of the *Internet*, and of *web-based service* approaches, tools, and applications, has greatly influenced and modified the whole IT arena. The second boost was initiated by *mobile technology* and the miniaturization of hardware components. Therefore, the paradigms of GIS have also changed, and the architecture of a GIS nowadays is quite different from what it was a few years ago (Fig. 6.4). Often, a normal browser can perform GIS tasks, especially simple GIS functions such as querying, displaying, zooming, and panning. Data are no longer restricted to the user's primary domain of interest and control, but they can in principle be imported from everywhere, anytime, and to any device. So, as in the early days, databases are still needed in current GI applications, but they can be distributed on a number of servers that reside anywhere in the cloud.

As the availability of GI steadily increases, we arrive at the term *ubiquitous geographic/geospatial information.* More and more real-world objects and phenomena are being mapped onto their digital GI counterparts. Also, web accessibility is getting better and faster, for the benefit of the mass market of mobile devices, where maps, satellite images, positioning, routing services, and even 3-D simulations are gaining ever-greater relevance. (For location-based services (LBSs) on mobile devices, see Chap. 22.) Also, in contrast to earlier times when a GIS consisted of a well-balanced combination of hardware, software, and data components, this setup is no longer valid. Hardware is exchangeable, data are ubiquitous, and for software, the borderline between the functionality of an ordinary Internet browser and GIS functionality is often fuzzy. Content is displayed that may, via web services such as Web Map Services (WMSs), be composed in an ad-hoc mode from different sources but have the appearance of a combined dataset, whereas in many cases the data themselves are not transported but are rather visualized on the fly. Web Map Tile Services (WMTSs), which are widely used in geobrowsers, can handle large preprocessed tiled chunks of images to support fast display of landscapes even on small mobile devices. The data remain at their various home localities, which is an asset as far as currency and lack of redundancy are concerned.

The question of whether a tablet or a smartphone with a number of georelevant apps installed and with data residing in the cloud can be seen as a GIS is a rather theoretical one and certainly does not interest the general public. However, these new developments have opened up new user segments. They do not replace traditional GIS users: those in the domains of public administration, utility companies, and communities in the diverse domains of science that are deeply rooted in models of a geospatial nature. However, these new web-based and mobile amenities have become important additions to the core elements of GIS technology, especially since they lend themselves to more user-friendly handling that is less dependent on office hours and strict work protocols. Being able to work from home, to perform some of the job requirements in different places, or to integrate the strengths of other experts or other systems via web services greatly enhances the value of geographic information and the associated technology as a whole.

Another contribution to GIS—one that is less spectacular than the web-based and mobile aspects but still very important—was introduced with *object-oriented modeling* and the corresponding methods in programming and database structures. Object-oriented methods bridge the gap between data and functions via the definition phase prior to the insertion of data. For each class of objects, an individual and particularly suitable function to use for data capture, storage, and retrieval may be defined. This relieves the burden on the functional shell that is built around the core database. One step in this direction has already been explained in previous paragraphs. Object-relational database management systems relieve the outer shell of many typical definitions of data types needed in geospatial applications, as well as many typical operations that conform to such data types. However, they are still based on the relational concept of tables, with rows (for the objects) and columns (for the attributes of such objects). In addition to the pure relational elements of a table, a geometry column is used, which takes care of the geospatial aspects mentioned before. Object-oriented databases go one decisive step further: they are no longer based on relational tables. Instead, object classes take on the role of the principal building blocks, adhering to the principles of data abstraction, encapsulation, modularity, polymorphism, and inheritance. Object-oriented GISs are yet to gain wide acceptance in the arena of GIS products; however, they carry great promise, since GISs—in contrast to many other information technology domains—characterize structures that are often more complicated and less unified. As an example, let us look at the way a building, as an object, can be constructed of many different components (base area, walls, roof) that can be further decomposed into still simpler objects. Each level of this decomposition can be seen as a different object class. In contrast to this GIS domain, the books in a library constitute a rather flat structure with components that are uniform. It would be easier for an information system dealing with entities from this domain to adopt a strictly relational database management approach.

A substantial change in the GIS paradigm was brought about by letting users not only consume but also produce geospatial information, making them *prosumer*s. This concept, called *voluntary geographic information (VGI)*, was introduced in [9]. It abandons the one-directional flow of information and frees the users from their passive role, making the whole process of creating and using GI more democratic, while also introducing many new questions about the quality and sustainability of GI. *OpenStreetMap* (OSM, www.osm.org) is the best-known example of VGI. Users are invited to contribute street geometries and many other features to an ever-growing database of geographic content. Clearly, such data add a flavor that is not present in authoritative data, but they cannot provide the same level of consistency and currency as the latter. However, in response to OSM and related initiatives, public authorities have opened up some geodata for almost unrestrained use by anyone under the label *Open Government Data* (OGD).

Together with VGI, the rapid increase in social media activities has had an impact on GISs. The fact that, in principle, any user may contribute almost any content to a given GIS necessitates the introduction of databases that can handle huge amounts of data that do not follow a predefined structure. *Key-value* or *graph databases* are examples. *Linked Open Data* and the *Semantic Web* have become important issues in this enlarged world of GISs. "It is about making links, so that a person or machine can explore the web of data. If you have some of the data, you can find other, related, data." This 2009 citation comes from Tim Berners-Lee, a principal originator of the idea of Linked Open Data, who is also known for having coined the term *World Wide Web* in the early days of the Internet. He writes about motives and design issues for Linked Data at https://www.w3.org/DesignIssues/LinkedData.html. This "birds of a feather flock together" aspect inherent to Linked Data as well as social media can be used for the benefit of GI applications. (Like any other human inventions, it also has drawbacks; for instance, in social media, we have lately become aware of the danger of getting deeper and deeper into a *filter bubble* of "more of the same" ideas, convictions, views on life, etc.) ISO has created a reference model for *Ubiquitous Public Access* (UPA) in its standard ISO 19154:2014, bringing traditional GI standards together under one roof with location-based services (LBSs), Linked Open Data, and ubiquitous computing environments.

Returning to GISs for a final aspect that is fairly new, the *Internet of Things* (IoT) has entered the arena, providing additional data originating from sensors that can be attached to almost anything one can imagine. This is not new to us. Any middle-class car bought nowadays is equipped with scores of sensors attached to the wheels, brakes, fuel tanks, and door locks, as well as sensors for light, temperature, road obstructions, conflicting vehicles, sudden changes of lane, etc.—sensors that make driving safer and easier. Likewise, IoT concepts have also entered the domain of GIS analysis. As an example, a routing application may provide different advice for route selection depending on the signals the system receives from IoT beacons registering congestion, snowfall, or accidents. Similar benefits can be seen when dealing with meteorological extremes (high/low temperatures, heavy rainfall) and utilizing sensors for early warning and monitoring. New chances and risks stemming from the IoT concept are being discussed widely in conferences and web groups.

To sum up, the modifications that have in recent years been transforming classical GISs into newer forms of geospatial analysis tools are as follows:

- *Web-based and service-oriented approaches* have led to a client–server architecture
- *Mobile technology* has made handheld devices with GISs ubiquitous, opening up a whole new market
- *Object-oriented concepts* have been partly introduced into GISs via object-relational databases
- *Voluntary GI* makes information flow multidirectional and assigns a more active role to users
- *Social media* has entered the picture, enriching the content, but calling for intense quality checks
- *Linked Open Data* significantly widen the range of information that is supplied in a GIS
- The *Internet of Things* can often substantially improve the significance of a GIS analysis
- *Ubiquitous Public Access* provides a common framework for traditional and new GI approaches
- Traditional GISs also benefit from *mobile and service-oriented technology.*

6.2 GIS Functionality

6.2.1 Categories

The list of GIS functionality categories can be arranged in many ways. There is no ideal arrangement, since we eventually have to arrive at a sequential setup, whereas many functions interact with each other in a manner that corresponds to a network rather than to a sequential list. However, a list is a way of ordering chunks of information that is easy for our minds to grasp. Our simple list follows the lifecycle of data from creation via structuring and storage to analysis and presentation. Each GIS contains representatives from each of these groups, although the quantity and quality of the functions in each group depend on the respective application range. In fact, GISs can be categorized by the degree of importance they allot to each group, giving rise to categories such as:

- data capture systems,
- administration systems,
- analysis systems,
- presentation systems.

A *data capture system* will put the greatest emphasis on the first phase in the lifecycle of geographic data. As an example, let us consider a topographic mapping authority in any given country that is in the process of converting large quantities of paper maps into digital form, and not only into a graphical pixel-based form such as an electronic image. For instance, road features are to be extracted from this image, and their geometries and several semantic attributes are to be stored in the database. Such a job usually requires semiautomated preprocessing, including scanning and image processing, and subsequent interactive editing. Even though data capture is its prime purpose, this system will also need a certain amount of administration for the data that have been captured. It will need some analysis functionality (for example, to find out whether all road segments do indeed connect in a topological way and, if they do not, to create the topology). Certainly, it will also need some presentation and visualization tools to assist the operators in their jobs, giving them feedback on what has already been done and what still needs to be done. So, these other three categories of administration, analysis, and presentation play a role, albeit minor ones compared with the role of data capture.

An *administration system* puts the main focus on the long-term storage of data, keeping them consistent and up to date. Examples include land information systems with cadastres that have to keep such data available for years and decades. They administer the data. Occasionally, though, some minor data capture is necessary for updates. Analysis functions and presentation functions help with the maintenance of such administrative systems. So, again, here we have a main focus and three subordinate interests.

In an *analysis system*, the situation is different. These systems put the main emphasis on one or several of the types of analysis that are discussed in the following sections; for example, a utility company providing power to households throughout a province may periodically need to analyze the total power consumption, the statistical distribution of peaks in consumption, the typical breakdown rate of power supplies during storms, the capacities of long-range connections, the overall flow capacity of a network, and other parameters. Also, such an analysis system needs some tools from the data capture, administrative, and presentation domains.

Finally, a *presentation system*—as we know from the many Internet applications that provides only the final graphical visualization of a given setup—conforms to what has been said about the four categories of systems. Certainly, normal Internet users cannot interact with it any more than zooming in or out, panning, and selecting and deselecting layers. However, behind the web barrier, such a system must also be able to capture some data that are needed to edit geospatial data, and it needs administrative and analysis functions to a certain extent.

6.2.2 Data Capture Functions

Chapter 9 is exclusively devoted to data capture in its different forms. For each category of capture techniques, it describes the requirements, the current state of the art, and the results. Also, data capture crucially depends on the model of reality to be attained. Modeling is extensively discussed in Chap. 1. Therefore, at this point, we can restrict considerations to the conceptual characteristics pertaining to more or less all data capture techniques, as well as their impact on the essence of a GIS and on the potential of such a system.

Bringing geospatial data into a GIS is one of the most challenging tasks among all the functional categories explained in this section. It has to encompass a large number of capture technologies, such as:

- Global satellite positioning systems (for example GPS, Galileo) assisted by inertial systems
- Geodetic surveying in all its different forms
- Laser scanning (airborne, terrestrial)
- Photogrammetry and satellite-based remote sensing
- Any combination of the above technologies, supported by cellular network and indoor positioning, a range of mobile mapping techniques, and UASs (unmanned aerial systems).

The quality of all subsequent processes to which these data are subjected depends to a great extent on the amount of care and attention paid to the capture process. To put it plainly, the output cannot be better than the original input. No matter how clever and intricate the subsequent editing functions, the transformation and structuring functions, or the analysis and visualization functions, they cannot undo a careless or sloppy way of obtaining the primary input. So, while all the other functions described in this section can, in principle, be redone if it is decided that they do not work in the intended way or they produce poor results, this easy way out does not apply to data capture, for several reasons. First, it is one of the most costly tasks in the lifecycle of data, involving a large amount of manpower, organization, and preplanning. Therefore, in many cases, the data are acquired once and for all. This is especially valid for data captured by surveying methods—be them traditional methods, such as a theodolite, or satellite-based methods. In any case, fieldwork is necessary. Also, the original setup may be lost. Consider a remote sensing campaign that provides satellite imagery or airborne techniques. The utility of such capture methods depends heavily on the season because of foliage, droughts, precipitation, or weather conditions. Also, since data may already be in use by several applications, it may not be possible to simply renew the data without risking those applications.

Another aspect that pertains to data capture in general is the fact that it can be seen as a process that maps the real world to a digital representation of the real world (Fig. 6.5). Often, it is argued that it is not the whole real world that is mapped, but only a *Universe of Discourse* (UoD)—the subset of the real world around us that corresponds to a specific application. As an example, in land cadastre applications, we talk about parcels, usage, property, survey points, etc., while we talk about start and end points, via points, routes, road segments, intersections, toll roads, traffic regulations, etc., in route planning applications. A third example is given by a municipal authority that administers the situation, infrastructure, and makeup of streets (e. g., the number of lanes, the sidewalks, lampposts, gutters, paving, and sewage sys-

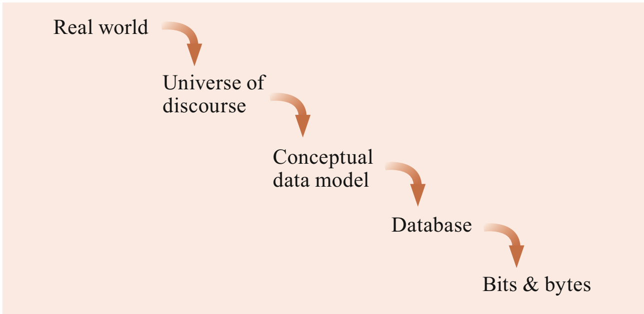


Fig. 6.5 Modeling

tem serving this part of the street). This example shows that any street in the real world may be considered in a multitude of UoDs, since it can be seen as a parcel of public property (i. e., an area), a connection in a traffic network (i. e., a line), or a combination of individual features that, when combined, make up the publicly accessible spaces in a city. At this point in the discussion, it becomes apparent that each UoD eventually needs to be mapped to a different *conceptual data model*, leading to different data structures in a GIS. This, in turn, tells us that there is no such thing as a universally applicable data model. Instead, each application requires its own data model, its own data structures, its own geometry, its own semantics, and its own accuracy requirements. Of course, this will not be practical nor financially feasible, and there is the problem that many future applications may not be foreseeable at the time when the data are captured. So, in practice, a compromise must be reached between theoretical prerequisites and practical restrictions. However, it is important to stress the fact that, before it can be successfully used, virtually every application needs a lot of special consideration with respect to the data available and the techniques used during the capture phase. This, by the way, is an essential ingredient in all arguments about the importance of *metadata* (Chap. 14), since metadata can be used to render—among other things—a detailed *report* on the lifecycle of data, especially the conditions under which they were born (i. e., captured).

Data capture will create digital data constituting a mapping or partial replica of an application-dependent subset of the real world. It cannot be a replica in the strict sense; rather, it is an *approximation to reality*. The real world is characterized by an infinite number of aspects, only a finite number of which can be taken into account. The quality of the resulting data will greatly depend on the way this subset is chosen. This pertains to all aspects relevant to the information involved.

For example, let us consider a mountain road that is to be mapped into GIS data (Fig. 6.6). Apart from the fact that an information system database is finite, several other choices need to be made. The data geometry will, in most cases, be 2.5-dimensional, observing the fact that north and east have far more importance than height, even for a mountain road. While the first two dimensions are mapped to a two-dimensional (2-D) coordinate system, the height is often carried as an attribute. This disparity between the importance of the first two dimensions and the third one is signified by the value 2.5. Of course, such a strategy can only work if any given line in the resulting geometry has, for each north and east value, only one height value, so there is a unique and invertible mapping between the 3-D surface and the 2-D plane. Such an assumption makes sense for mountain roads, but it does not make sense for subway lines, for example.

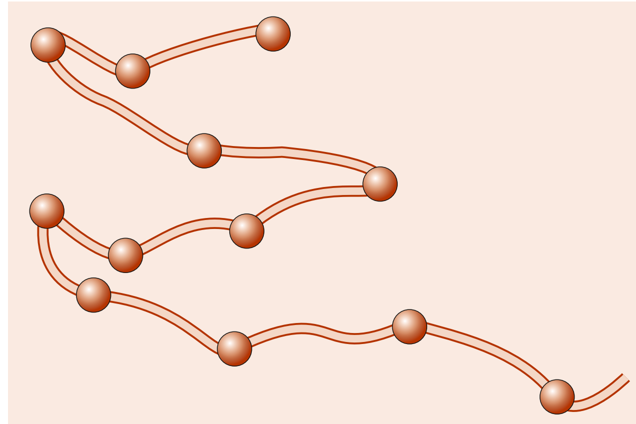


Fig. 6.6 Modeling the geometry of a mountain road

Having taken care of the third dimension, we still have to decide upon the remaining two dimensions. A road is a general surface in reality, but we have now reduced it to a plane surface. In most cases, the modeling process will further reduce it to a line—the center line or axis of the road, with the second dimension stored in another attribute corresponding to the width of the road at certain points. So, the second dimension has the same fate as the third dimension before: it ends up as attribute values, leaving the geometry of the mountain road as a one-dimensional line. Such methods can be seen as simplifications of the real world. The art of modeling consists of simplifying without overdoing this. The last simplification step would be the selection of certain points along the road axis to represent the road in its digital replica. It is impossible to choose infinitely many points of course, and it is not recommendable to choose too many points (because this would make the data volume explode) or to go to the other extreme (if there are not enough points, information on the presence or absence of turns—a very important characteristic of a mountain road—will not be available). So, again, the art of modeling and the art of capturing just the right amount of data at just the right places will determine the success of any further applications built on that dataset.

The previous discussion provides a short insight only. However, it can easily be transferred to other domains such as sensors in photogrammetry and remote sensing, other airborne and spaceborne sensors, and mobile mapping systems in general.

To sum up, any data capture technique must, to obtain the maximum benefit of the data in a GIS, adhere to the following principles:

- *Suitability* for given or envisioned applications;
- *Reducing the richness of the UoD* by performing the appropriate amount of simplification;
- Observing *best-practice quality criteria*.

6.2.3 Update Functions

The creation of data and their insertion into a GIS will often be followed by amendments, corrections, updates, and deletions. Almost any information system, and therefore also a GIS, will be created with long-term usage in mind. The data capture process is characterized by the requirement for high *geometric and semantic quality* as well as *stable and suitable structures*. This, for example, includes the structures used for features and for topological networks. See the previous section for more details on data capture and the following section for more details on structuring. The consequence of all this is the well-known fact that data capture is, in general, a costly task. Costs arise partly due to the sheer volume of data and partly because of time-consuming capture tasks that involve a considerable amount of manpower. If data are found to be erroneous, ill-structured, or incomplete, it is seldom advisable to capture them anew; rather, they should be subjected to amendment and update procedures.

This sounds simple, but if those amendments are done carefully and in a *user-friendly* way, this task may turn out to be quite difficult. Any primary input is far easier to handle than subsequent amendment processes. In the initial data capture phase, data usually arrive en bloc in the database. Consider as an example the insertion of data that model road geometries (Fig. 6.7), since we used this example in the previous section. As outlined there, in most cases, the geometry of a road segment from one intersection to the next will consist of a sequence of points denoting the axis of the road, while the road width at each of these points is modeled by an appropriate attribute value. The height value is handled in the same way. This procedure is straightforward if the whole road segment is inserted into the database in this way. Now consider the case where an additional point has to be inserted, or one point has to be eliminated from the sequence. While the user knows what has to be done, this cannot be so easily achieved by the system, because any insertion in the middle of the sequence has to be handled differently from an insertion at the beginning or at the end. The problem can be overcome, but the challenge is far bigger, since the user-interface shell of the GIS must reckon with a score of different situations that can arise due to erroneous inputs.

While such questions arise at the microscale, i. e., for individual inserted or updated features, there is also a macroscale problem that has to be analyzed so that, eventually, ways can be found to deal with it. GISs have been designed to provide a long-term basis for geospatial analysis. GIS data will have a long lifespan. Cadastral data are a prominent example, but many other geospatial data also describe objects or phenomena in the real world that have existed for a long time and whose existence is likely to continue into the future. These long lifespans of all the individual GIS features did not all start at a common *Big Bang*. Rather, their lifecycles meet and overlap on the real-world axis of time in an unforeseeable manner. Additionally, we have to observe yet another time axis corresponding to the time that the GIS data replica of a real-world object or phenomenon was inserted into the GIS. To add yet more complexity to the problem, we also have to consider the fact that GIS data originate from many different sources at different times. This is a great bonus and strength of any GIS, but also one of the main causes of problems arising during the update process due to incompatibility issues. When different data collections from different data sources have to be merged, this can be seen as an update process at the macroscale.

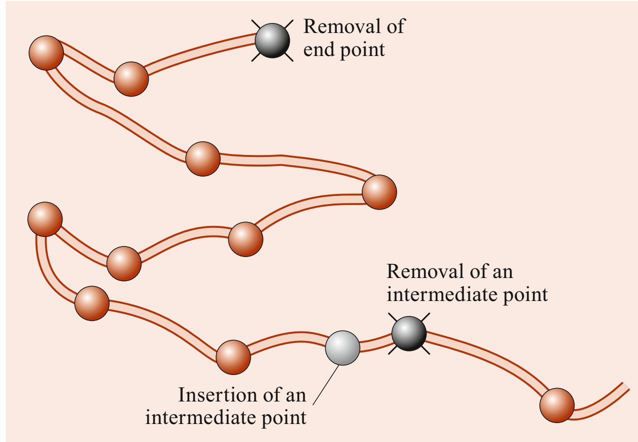


Fig. 6.7 Updating the geometry and topology

All the aspects mentioned above will add to the complexity of GIS update functions. Let us again make use of the road example. This road may be part of a provincial road network that has been digitized at time point A and has been present in the GIS database since then. This road network has been checked for consistency in both the geometric and topological senses. When another layer of GIS data containing roads that belong to another administrative hierarchy (e. g., a road maintained by a local township administration) is added to the system, it is likely that the combined network of provincial and local roads will have a topology that is more detailed and will typically contain more intersections (topologically speaking, these are called *nodes*) than the provincial network. An update function in this general sense, which merges two datasets that each have their own special lifecycle, quality measures, and application domains, thus creating a new combined dataset where all these criteria need to be harmonized, requires quite a bit of effort in both the database shell of the GIS as well as the human–computer interface shell.

To sum up, update functions can be characterized in the following way:

- They are much more complex than the original input functions (this is true of the database shell, but even more so for the user interface shell);

- Update functions cover geometric, semantic, and structural corrections at the microscale (for any feature);
- They also cover the macroscale when different datasets need to be merged;
- Different timescales (real-world and GIS insertion times) complicate the process further.

6.2.4 Structuring Functions

Data enter a GIS in a more or less basic structure: as points and lines for vector-based applications and as pixels and images for raster-based applications. Such basic data are the building blocks for more complex structures that are needed to model the real world in a way that can be put to use for typical GIS applications. These functions assemble basic entities into more complex structures. Let us again use the illustrative example of road geometries. Starting with individual points that are captured by GPS methods, by digitizing points on the screen, or by some other method, it is necessary to connect two adjacent points, forming a line. This may seem self-evident at first glance, but it is not. Let us explain the difficulty by looking up at a starry sky on a cloudless night. What we see—the stars—can be considered points. Forming clusters that correspond to signs of the zodiac and then connecting the stars in each sign in the appropriate order is not an easy task. If the structuring process is performed in even a semiautomated way, the difficulty becomes even greater (by several orders of magnitude). Similar things can be said for the interactive and semiautomated structuring of GIS data. It takes a lot of experience and a concise *set of rules*—some of them pragmatic or based on experience only—to make this process work. The order of the points forming the axis of our example road can, in many cases, only be guessed at, unless we have additional information that can be used, for example, a digital terrain model. Ambiguities may be avoided by using some assumptions that make sense; for example, a road is more likely to follow the terrain than to go up and down in a zigzag manner. Often, a decision can also be made easier if we consider *topological rules*: a road may not cross itself; a figure-of-eight shape is usually considered to be a topological error; undershoots and overshoots need to be corrected; ambiguities for nearby points need to be resolved by averaging them; etc. Structural deficiencies have to be eliminated (Figs. 6.8, 6.9).

Once the sequence has been determined, the whole road can be structured, like forming a string of pearls. What has not been decided yet is the geometric form of the connecting lines. They may be straight lines or curves in the real world. In the case of a mountain road, curves are more likely. If we consider streets in a US suburb, straight lines are more likely. If we create a model for rivers instead of roads, straight lines between adjacent points are almost unthinkable. This shows the vast range of considerations that have to be included in any structuring process.

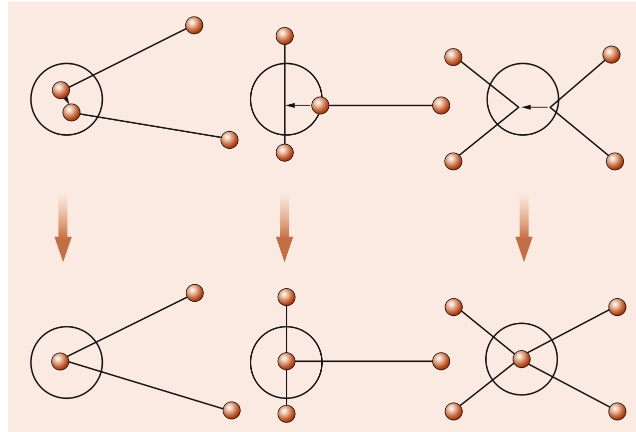


Fig. 6.8 Structural deficiencies in a linear topology and how to amend them

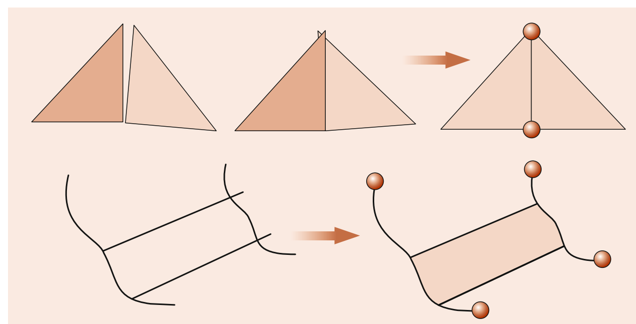


Fig. 6.9 Structural deficiencies in an area topology and how to amend them

In a similar way to the one-dimensional structuring of roads, rivers, etc., two-dimensional (2-D) structures may be formed from either points or lines. If the points from our starry sky example denote the corners of buildings that should be modeled in a map-like two-dimensional way (i. e., neglecting heights, storeys, etc.), it will be necessary to arrange the corner points for each building in a closed ring sequence. The simplest form of a building in this two-dimensional projection will be a rectangle. Again, there will be rules that have to be observed, as in the previous case, where we did not allow figure-of-eight shapes. We also have to consider other topological rules (adjacency, no overlapping, etc.), and there will again be restrictions based on experience. As an example, we can rule out buildings with an extremely elongated shape or buildings that would have an area of negligible size.

With each dimension added, the complexity of the structuring increases. Three-dimensional structuring, as performed, for example, in 3-D city models, represents the most complex form of geometric structuring. We can begin the structuring starting from points in 3-D space originating from

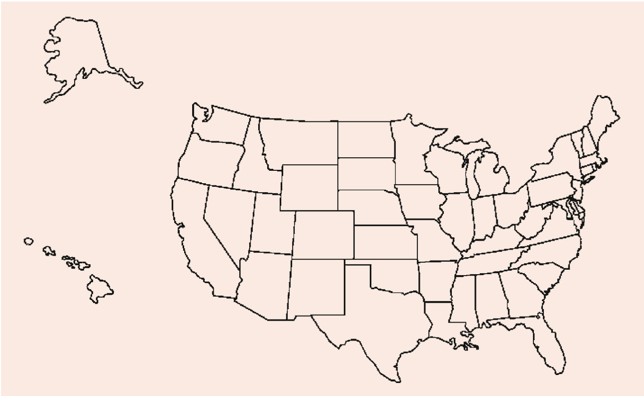


Fig. 6.10 Structuring according to semantic criteria (US states)

surveying or laser scanning. Those points will typically be the ground corner points as well as characteristic points of building fronts and roofs. Points can be structured into lines, for example, a roof line or a base line. Lines may again be used to form a two-dimensional structure such as a house front or part of the roof. Then all these components can finally be assembled into a 3-D structure. This structuring follows an approach that considers only the outer cover of the building. There are several alternatives. For more on 3-D city modeling, please refer to Chap. 21.

Structuring is not restricted to geometric and topological aspects. The *semantics* as well as *time* will often give rise to structuring. An administrative area that is built from several topologically independent patches is one example, as is the mainland of the USA together with Alaska and the Hawaiian Islands (Fig. 6.10).

So far, we have only discussed the structuring of vector-type GIS data. Considerations for raster GIS data are just as important. Let us, for example, consider an image stemming from an airborne campaign, and let us assume that this photo has been subjected to rectification processes, yielding an orthophoto. The human eye can differentiate visible features such as rivers, roads, and built-up areas in this photo. If we now let a GIS structuring function attempt to do this in at least a semiautomated way, many of the abovementioned questions also arise in this context. Grouping the pixels of the photo into clusters that belong together can be seen as a classification process. There are many pixel-based and object-based methods and tools to perform such structuring (Sect. 10.5).

To sum up, structuring functions can be characterized as follows.

- They *assemble basic components* stemming from data capture into high-level application-friendly structures
- *Geometry* and *topology* as well as *semantics* and *time* give rise to a need for structuring
- Structuring must adhere to *application-dependent rules* and often to *informal knowledge* (experience).

6.2.5 Transformation Functions

The term *transformation* in a GIS environment is widely associated with *coordinate transformations* (Fig. 6.11). An essential benefit of using GIS is the opportunity to bring many different types of geospatial data together under one roof, enabling them to be compared with respect to identical, overlapping, or nearby locations so that conclusions can be drawn about interactions between geospatial aspects. Bringing together different types of data is only possible if the spatial references of all data are resolved so that they can be transformed into a common frame. This signifies that coordinate transformations are really at the core of any GIS. Such processes are generally triggered implicitly, without waiting for an explicit command from the user. However, in order to be able to discuss the benefits and risks of comparing or even combining disparate datasets, it is necessary to acquire a basic understanding of coordinate transformations and the reference systems and reference frames on which they are built. Geodesy deals with this, and consequently it is discussed in detail in Chap. 8. Coordinate transformations are necessary not only in the inner core of a system, where they largely go unnoticed by the user unless they fail to work properly; at the user interface level too, any zoom or pan involves aspects of coordinate transformations, and therefore they should be discussed in a comprehensive list of GIS functions, even if such zoom and pan functions are considered very basic and hardly worth mentioning. However, it is not difficult to find an example where zooming becomes slightly more complicated: large-scale maps are typically presented in a plane coordinate system that is rectangular without losing too much accuracy. In contrast, regional or state maps usually have to take the Earth's curvature into account if they are not to look awkward. So, it can very easily happen that GIS users notice the relevance of choosing a coordinate reference system that best suits the current zoom.

Coordinate transformations are not the only GIS processes that are modeled by transformation functions. Another important representative of such functions is the *conversion* between *raster data* and *vector data* (Fig. 6.12). These two modeling strategies that yield two different types of data structures have in the past often led to two different, and to some extent incompatible, worlds. Often, a GIS was seen mainly through the vector-type lens, with the raster world only present as a backdrop image. In recent times this has changed dramatically. Not only is raster imagery becoming increasingly

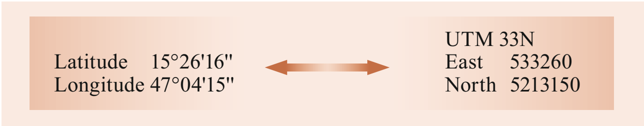


Fig. 6.11 Coordinate transformation

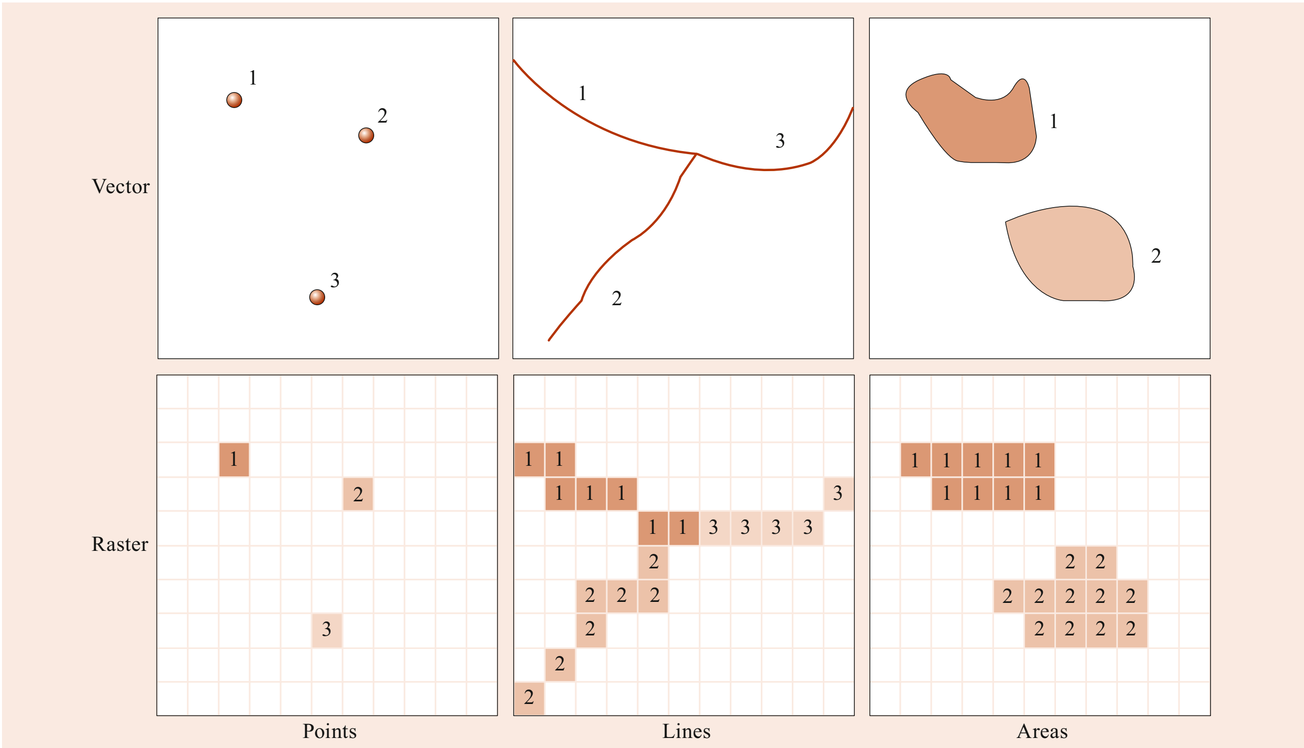


Fig. 6.12 Vector and raster data

available and—despite its volume—manageable, but a large number of image-processing functions can be put to good use for GIS data. Let us cite just one example. Cartographic generalization is one of the key issues in the area of visualization. Image-processing techniques lend themselves easily to generalizations based on raster data. It is therefore desirable to—at least temporarily—transform vector data into their raster counterparts. Of course, the inverse techniques are also needed. We will not go into detail here, because all of this has been successfully used for many years in image processing. The reader can find ample coverage in the literature; for example, *Pavlidis* [10], *Pratt* [11], and *Gonzalez and Woods* [12].

In this section dealing with transformation functions, it must be noted that other GIS functionalities can also enter the discussion. These include interpolation and approximation functions and adjustment theory, computer-aided design (CAD) construction functions, as well as transformations in image and raster data processing (Chap. 2).

To sum up, transformation functions can be characterized in the following way:

- In the form of *coordinate transformations*, they are ubiquitously present in GISs and are of general importance
- *Transformations between vector- and raster-type data* are increasingly taken from image processing
- Tools from *mathematics, approximation and interpolation theory, and adjustment theory* can be used.

6.2.6 Storage, Checking, Archiving, and Data Transfer Functions

In Sect. 6.1, we defined information as an answer to a (sometimes implicitly posed) question that heightens the level of understanding of the inquirer and/or makes it easier for them to achieve a certain goal. For a GIS, the metaphor of an engine or a vehicle that brings us nearer to the solution of a geospatial problem is therefore appropriate. Any vehicle needs an infrastructure it can run on, for example, roads, railroad tracks, waterways, etc. For a *geospatial problem solution engine*, geographic data provide such an infrastructure. Roads, railroads, and waterways are essential for the functioning of modern society and therefore considerable investment is directed into maintaining them. Likewise, a *geospatial data infrastructure* needs to be well maintained since it provides a basis for current and future GIS technologies and for *geospatial solution engines* in a wide range of potential implementations, and is therefore of considerable value. Effort, time, and expense have been invested in creating it (Chaps. 15 and 30). It will have to be available for a long time to come, so proper maintenance is necessary. There are scores of functions to make data fit for long-time storage, to perform lifelong checking of consistency and other quality parameters, and to facilitate the interoperable use of geographic data, including transfer and services. The metaphor we have invoked has in recent times become widely used due to initiatives for geographic data

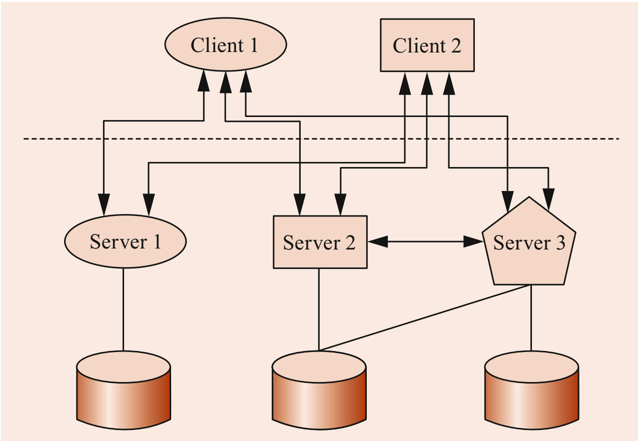


Fig. 6.13 Interoperability ensuring standardized access to different servers

infrastructure at global (Global Spatial Data Infrastructure, GSDI), national (National Spatial Data Infrastructure, NSDI), and regional levels. The European Infrastructure for Spatial Information in Europe (INSPIRE) initiative, which is evolving into a framework of harmonized national laws supporting *interoperability* (Fig. 6.13), can be seen as falling in between the global and national levels. More information on geospatial data infrastructure and interoperability can be found in Chaps. 15 and 17.

The range of GIS functions that should be mentioned in this context starts with all database operations that are available at the user interface level. Database insertions, updates, and deletions are often implicitly invoked without needing an explicit command from the user. This is the case, for example, when data capture or update functions are executed. When users manipulate geographic data on the screen, they can safely assume that all their actions at the user interface level are mirrored by appropriate functions at the lower level of database management. However, there are cases where users explicitly need to invoke storage and archiving functions. This is comparable to a desktop office environment where the current state of progress is periodically archived but the user may want to explicitly trigger a save or even an archiving function at certain stages of the process. While the former simply saves the current state in order to be able to document it or to safeguard against system failures, the latter provides a means to keep track of several states in time—a *time machine* as it is called in some systems. Geospatial applications are often characterized by the need to keep track of past situations or situations that have become obsolete or historical. Think, for example, of cadastral applications, where not only the current ownership situation but also the history of a parcel and how it has evolved over time (including all partitioning and merging operations in the past) are of interest.

Another important category of functions to mention in the context of this section are those for *checking and validating*. Let us again use the metaphor that compares geographic data infrastructure to roads and railroads. Periodically, all of the aspects of such an infrastructure must be scrutinized; for example, the desired geometric quality, topological consistency, and semantic and temporal validity. Additionally, the structuring rules need to be checked for compliance. Such functions may include small corrections that should be done automatically or under user surveillance. Consider, for example, a regional directory of emergency services. Any new entries and also deletions and updates call for interaction with the authority responsible for keeping and providing the directory. Changes in some attributes, such as the update of a telephone number, can be traced automatically and therefore also be corrected without needing interaction.

This leads us to the last group of functions that need to be discussed in the context of this section: those dealing with the *transfer of geographic data* and *services*. Historically, the transfer of geographic data from one system to another was one of the main challenges in GIS technology. The need for interface standards spawned interdisciplinary and international standardization initiatives, for example, those at the CEN and ISO levels. CEN/TC 287 *Geographic information* (https://standards.cen.eu) and ISO/TC 211 *Geographic information/geomatics* (https://committee.iso.org/home/tc211) paved the way for many advances in standardization. The Open Geospatial Consortium (OGC, http://www.opengeospatial.org) put the term *interoperability* on its banner. The ISO/CEN and OGC developments together provide some essential steps beyond the simple data exchange paradigm of the past. Interoperability denotes the ability of different systems to work together and to make use of each other without interfering in internal procedures of the individual systems. This leads straight into the modern paradigms of GIS services that can be invoked on the Internet (Fig. 6.14). Traditional geographic data exchange is a special kind of service, but many other services (for example, WMS) function without needing to exchange basic geographic data. A route finder is an Internet-based service that is even better known to the public. The result of a route query is not retrieved by transferring geographic data on roads to a local system and invoking a routing function there, but by transferring a map-like representation that can be visualized in a simple browser. For further reading on interoperability and GIS services, the reader is referred to Chaps. 15 and 17.

To sum up, data storage, archiving, checking, and transfer functions can be characterized in the following way.

- GIS data infrastructures with *long-term stability* call for best practices in data storage and checking
- *Archiving geographic data* meets the demand for time-machine-like capabilities of a GIS
- *Standards on data transfer and interoperability* pave the way for web-based service architectures.

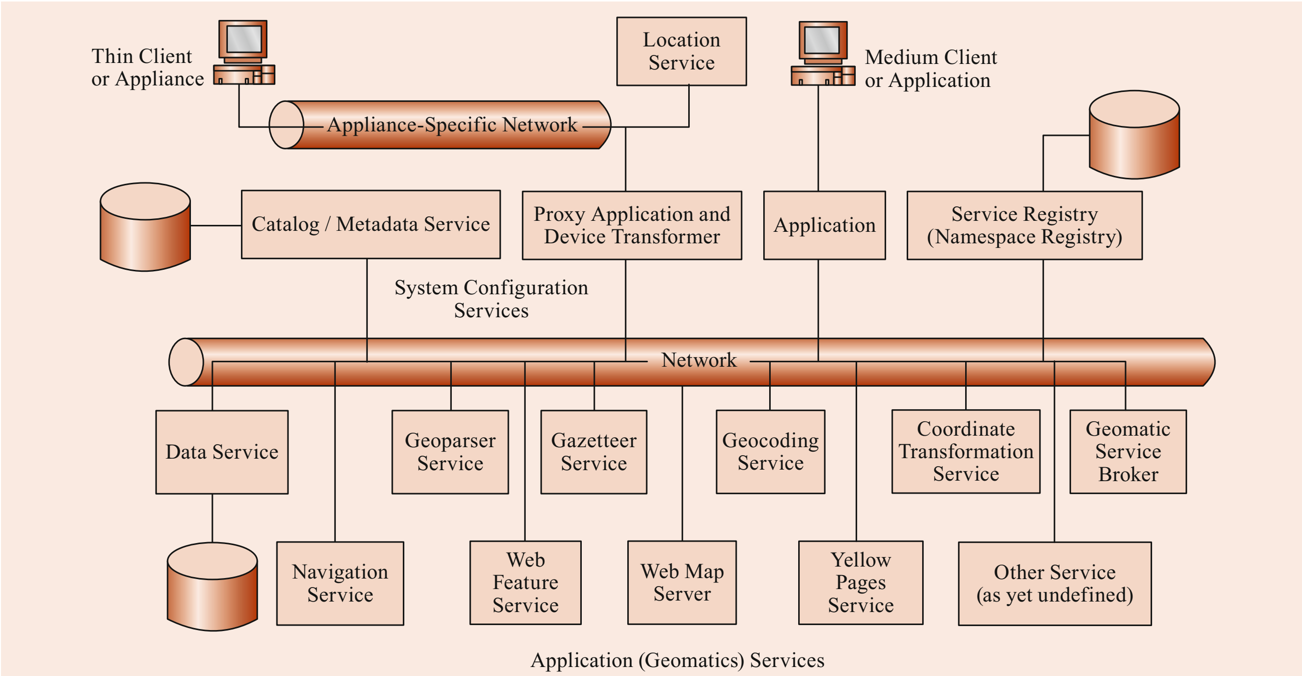


Fig. 6.14 Services according to ISO 19132 *Geographic information—Location-based services—Reference model*

6.2.7 Data Request and Retrieval Functions

Information is an answer to a—possibly implicitly posed—question, as was outlined in the first sections of this chapter. Web-based search engines are widely used information retrieval tools that provide answers to questions about content. For example, when searching the Internet for information about clock towers, we may retrieve an answer that a clock tower is a tower with one or more (often four) clock faces, and that it is usually part of a church or municipal building such as a townhall but can also be freestanding. If we seek an answer to the *what* but place equivalent importance on the *where*, geographic/geospatial information will result. We may find out about the Westminster clock tower in London, the Zytglogge in Berne (Switzerland), and the Uhrturm in Graz (Austria), as well as many other examples worldwide. It would be nice to see a world map pinpointing the locations of clock towers, or even to find examples that are close to home—provided that the search engine can (and should) be informed where home is. When planning a weekend trip, we may even want to search for clock towers that are within a 3 h drive, assuming average driving habits and average traffic flow. We soon get into the details of a GIS search and query and find that we need to use topological GIS analysis functions based on road networks and possibly route-planning facilities (Sect. 6.2.8). Knowing where things are in relation to our current location satisfies a basic need, since hardly any action we plan to take and hardly any decision we have to make is independent of location. We humans try to pinpoint ourselves as well as the points of interest to us in space.

Computation is often an essential part of an information request, as the example of driving distances and driving time shows. Searching for a coffee shop that is not farther away than 5 min by bicycle requires some computing, as does a query about sightseeing locations that are not farther than 5 km from a vacation trip route or a mountain bike track that stays within given limits as far as horizontal distances as well as summed vertical distances are concerned. All these examples show that GIS functions for request and retrieval could also be categorized as *query functions*. However, typically, such a geospatial search and query question and its answer are decomposed into two parts, with one operating at or near the user-interface level of the GIS (dealt with in this section) and one operating in the *analysis engine* of the GIS (dealt with in the next section). We first have to formulate our question, and this should not be done solely via character strings. Introducing search methods that are based on semantics as well as geospatial concepts brings considerable challenges. Searches based on ontology and Semantic Web concepts are becoming increasingly important. For example, an ontology-based search would not necessarily presuppose that the user will provide the string *clock tower*, but would instead use hints from the user, who may not have expertise in posing formalized questions. The search would, like a detective, combine several parts of the puzzle and let the final form of the search query boil down from concepts such as building structures, historical landmarks, sightseeing, medieval city centers, etc., and what they have in common. This is also a typical task of a knowledge engineer, leading to a *knowledge-based system*, an advanced type of information system (Chap. 17 on the geospatial Semantic Web).

The geospatial aspect of the question we have in mind adds yet another challenge. The question of *where* may be too fuzzy in terms of the capabilities of the system we intend to use. The geographic coordinates of any clock tower are of little use to the general public. A simple display on a map may be more informative, but if we really plan to visit the tower, data on the ways it can be accessed, the time it takes to get there, and the best route (depending on the means of transportation) are all necessary to give us the information needed for us to decide whether it is feasible and advisable to visit this sight. Here we use the term *information* in the sense described in Sect. 6.1—as an aid to help us make a decision. Up to now, we have discussed information retrieval for features that can be pinpointed, meaning that their locations roughly correspond to points on the Earth's surface. This holds for our example of the clock tower, even though such a tower is not a point in the geometric sense. However, its mapped replica in the GIS is a point feature that is tied to the location of its center point or its main entry. For line features, area features, and 3-D features, with each category adding yet another geometric dimension, things become more complicated. Evaluating the distance to a line feature such as a mountain road is more difficult than it would be for a point feature, even if we use the 2-D distance only and do not take into account any 3-D information or any obstacles on the way, such as a river or a steep slope that cannot be crossed. Projecting the current standpoint orthogonally onto all intermediate points along the road axis, computing the distance to its starting point and its end point, and finally choosing the road point with the minimum distance may answer the question satisfactorily. However, there may be ambiguities if we ask for the nearest road point. Adding yet another dimension and asking about distances between 2-D features (for example, asking for the distance between housing complexes with detailed geometries) will naturally make things even more complicated.

The distance between the Mississippi River and Colorado River is a question that cannot be answered satisfactorily. We could compute all possible distances from any point along the Mississippi River axis to any point along the Colorado River axis and choose the minimum, but adopting this procedure for the Mississippi River and the Missouri River would yield a distance of zero, since the two rivers join at one point. Is this a satisfactory answer? The same problem arises when we ask for the distance between America and Europe. Yet another degree of difficulty enters the discussion if the delimiters of spatial features are not precisely defined. Examples include the question of the distance between Burgundy and the Alps or that between the Rocky Mountains and the Great Plains. All these aspects influence the way we have to ask a question, and therefore the system capabilities that help us to ask the question in an efficient way, as well as the way the GIS answers the question. Also, for the latter, the GIS needs to be equipped with facilities that are far more intricate than a simple map output with pins and/or mileage numbers displayed on it. Due to the approximate character of typical GIS queries, methods and functions must be provided to assist the user in determining the appropriate answer from several choices.

Performance issues are also part of the discussion about GIS request and retrieval functions. GISs typically have to handle very large amounts of geographic data. Efficient search algorithms have to be used to enable fast access and retrieval, not only because of the large volume but also because of the imprecision and multidimensionality of query predicates that describe the query condition. The time a system takes to answer a question should roughly correspond to the difficulty the user allots to it, on an intuitive scale, and not specifically to the data volume stored in the system. For a problem where the user has to put greater effort into formulating the question and interpreting the answer, it is tolerable if the GIS query function needs more time. However, the size of the data volume stored in the system will not increase the user's patience. Care has to be taken that the answer does not exceed the timespan of the user's attention. If their attention is diverted due to excessive waiting, the usability of the GIS suffers greatly.

To sum up, data request and retrieval functions can be characterized in the following way.

- They have a *big impact on the user interface*
- *Semantic and geospatial aspects of the query* need to be formulated and to some extent formalized
- Likewise, the *resulting information must be structured* in a way that suits the user's needs
- *Performance criteria* are important.

6.2.8 Analysis Functions

This is by far the largest group of GIS functions, but also the hardest one to confine or categorize. The number of special GIS applications is growing fast. Many such applications appear as newly added features in IT domains that can be quite distant from the definition of a GIS in the strict sense. A typical example is the routing and guidance applications that appear in many web page presentations of companies and organizations as well as on mobile devices. Typical subtypes of GIS analysis functions deal with *geometry* (for example, proximity analysis, overlays, and intersections) and *topology* (for example, network analysis, as in routing and guidance). *Simulation and planning* is another important topic that introduces a wide variety of models and analytical requirements, ranging from urban planning and its effects on flood disaster management to agriculture and geomarketing simulation models. Another notable segment of GIS analysis is constituted by *statistical functions*, for example, those that implement ranking, regression, trend surfaces, prediction, and filtering. We will discuss a few line interpolation and surface interpolation techniques such as digital terrain models (DTMs) that are standard in GISs. For other statistical functions, we refer the reader to Chap. 2.

Data request and retrieval functions (discussed in the previous section) usually appear in combination with analysis functions but deal with the way a typical query is handled and supported by the user interface of the GIS. The current section highlights the different aspects of the *analysis engine* that operates below the user-interface level but interacts with it in a symbiotic mode.

Geometric Analysis

For geospatial data, the geometry of features is of prime interest. Many—if not all—GIS analysis functions have at their core essential aspects of geometry. The location of a feature is a basic ingredient of geospatial information, as is its location in comparison with those of other features. Two or more sets of features may have to be overlaid and their geometries intersected. Ranges, neighborhoods, and buffers will be evaluated. In many GIS applications, this metric aspect of geometry, which is closely connected to measuring and rendering distance measures and eventually coordinates, will have to be complemented by a topological analysis rendering network-like connections, adjacency, and containment (which are, conceptually speaking, independent of distances and coordinates). These topological aspects and the corresponding analysis tools will be discussed in the next section.

Two-dimensional *Euclidian geometry* serves as a first intuitive approach to geometry. This is built on the concept of an idealized flat surface, which serves as a good approximation for many purposes, including that of this section: the geometric analysis of geospatial information. Another assumption that is reasonable is an orthogonal coordinate system with axes pointing east and north. Extending this into 3-D *Euclidian geometry* is straightforward. When we move to still higher dimensions (for example, using time as the fourth dimension), our imagination finds this difficult to visualize, but mathematically it is possible and is implemented. Non-Euclidian geometries are not that far-fetched. Consider, for example, the surface of a sphere as an approximation to the Earth's surface. *In spherical geometry*, triangles can have angles that sum to greater than 180° (Fig. 6.15), which is not possible in Euclidian geometry. Just think of going north from your current position until you arrive at the North Pole, then going south again on another meridian, traveling the same distance as before, and finally moving along the parallel of latitude towards your original position. This results in a triangle with two angles of 90° at its base plus the angle at the North Pole. Spherical geometry is not the only alternative to Euclidian geometry—there are many others.

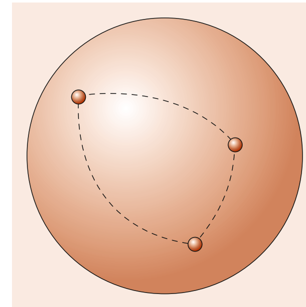


Fig. 6.15 Spherical triangle in which the sum of its angles is greater than 180°

In 2-D Euclidian geometry, the distance between two points is given by the square root of the sum of the squared coordinate differences. This is the famous formula of Pythagoras, since the distance line and the two coordinate difference lines form a right-angled triangle. The distance between a point and a line will be given by the length of the orthogonal projection of this point onto the line. The distance between a point and an area will be the minimum of all distances from the point to the boundary of the area. Line–line, line–area, and area–area distances are defined by appropriate extensions. Mathematically, a space that allows the measurement of distances between its elements is called a *metric space*. Any distance measure must conform to the rules of a metric space. The first rule says that a distance will always be greater than zero except for two elements that are equal, in which case the distance will be zero. The second rule says that the distance must be symmetrical, i. e., it is irrelevant whether you measure the distance from the first to the second element or vice versa. The third rule, called the triangle inequality, says that, if you do not go directly from one element to another but take a detour through a third element, the total distance will not be shorter than the direct path.

Euclidian space and the Euclidian metric seem to be straightforward (in both meanings of this word). However, in GIS, this often needs to be modified. While the *Manhattan metric* complies with the three rules postulated above, it takes another approach (Fig. 6.16). The distance between two street intersections in a city with a rectangular network of streets can be measured in the Euclidian way, but a more realistic measure in this case would be the number of blocks you have to wander east (or west) plus the number of blocks you have to go north (or south). The Manhattan metric is important in raster-based applications where the raster cells correspond to the city blocks in our example. However, for traffic and transport applications that are vector-based, this strategy is, in certain cases, more relevant to real-life requirements for vehicles and pedestrians than the Euclidian

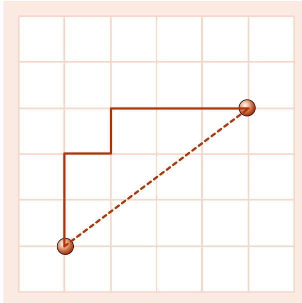


Fig. 6.16 The Euclidian metric (distance = 5 units) and the Manhattan metric (distance = 7 units)

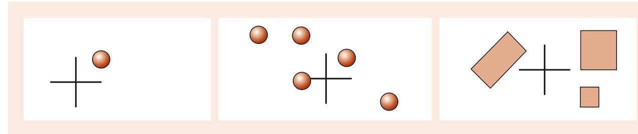


Fig. 6.17 Nearest-neighbor search for one or several points and for corresponding features

distance, which instead models a bird's flight from one block to another. There are several other extensions of and/or alternatives to the Euclidian metric that serve useful purposes in GISs. Using time as a distance measure may, in many cases, be more realistic than using mileage. Also, since geospatial data are always finite, we must generally allow for modifications, even when using the Euclidian metric. For example, the fact that line features are finite means that distances cannot always be realized by measuring orthogonal projections, since the projection points may be outside the range of the line feature.

Having defined the framework upon which geometrical analysis needs to be built, we can now proceed to some protagonists of such functions. A good start to this list would be a *nearest-neighbor search* (Fig. 6.17). Such a search sounds simple, so we might surmise that the answer is too. This problem arises in standard situations during interactive work. The user points the cursor at some feature on the screen because they have a question about the feature in mind or with the intention to use this feature in a subsequent analysis step. It is very difficult to point at a feature stored in the database precisely. A search must therefore be initiated, starting with the cursor position and retrieving candidates that are close to it. This search must deliver a result very quickly, no matter how large the database is. Any delay would be criticized, since users intuitively equate the tolerable waiting time with the difficulty of the question posed. This expectation is of course not correct, but such an argument is not acceptable either. Another problem may arise due to ambiguities, i. e., when the computation of distances to candidate points does not render a unique result. In this case, a spiral search of the nearest, second-to-nearest, etc., points is advisable, where the user is asked to decide. This is not difficult to achieve theoretically, but it puts more of a burden on a user-friendly interface. Finally, let us mention a conceptual aspect that often causes another kind of complication. Although geometric features are based on points, and higher-level structures are often identified by identifying their points first, sometimes users consider another approach to be more natural. Consider a map of the USA. When a user is asked to identify one state, they will place the cursor somewhere in the middle of a state and not at a point along its border (which would be ambiguous anyway). So, the appropriate user response depends on the given situation, and again a good user interface *should know* the appropriate action and reaction to a particular situation.

Polygon overlay and intersection is another protagonist of geometric analysis functions. Consider an application dealing with environmental protection zones for wildlife, vegetation, etc. The zones can be geometrically modeled by 2-D polygons consisting of points along the boundary of each polygon, where the points are connected by straight lines in a specified order. Typically, the polygon boundaries will follow natural boundaries, for example, topographic landmarks, rivers, or forest edges. In Fig. 6.18, area stamps in a simplified manner show their location and extent. These polygons should now be spatially compared with administrative areas such as counties or municipal districts in order to assess the area percentage of protection zones per administrative unit. Two meshes of polygons have to be overlaid, the intersection points have to be found, and their coordinates have to be computed. The resulting mesh should have all the original points plus the new intersection points, and the new tessellation will often be finer than either of the original ones, but it will never be coarser. The new intersection polygons must again be correctly formed by putting their boundary points in the appropriate order. The intersection polygons answer the question of where environmental protection zones could interfere with the agenda of a municipality, and which municipality has to find a balance with which protection zone. Also, areas can be computed in both km^2 and percentages, and this can be done for each individual administrative unit, for each individual protection zone, or in total for the whole set of data.

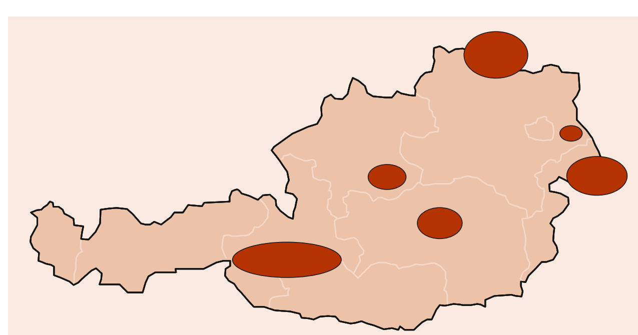


Fig. 6.18 Polygon overlay and intersection

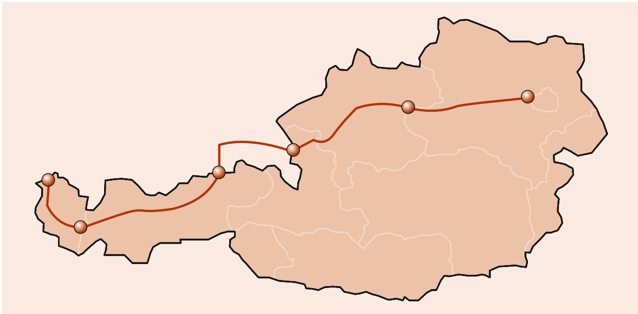


Fig. 6.19 Overlaying lines on polygons

A similar approach is used to *overlay lines on polygons* (Fig. 6.19). Consider the case where a road network is to be overlaid on the protection zones. The questions are: which road segments fall into which protection zones, where do those road segments enter a zone and where do they leave it, how long are those road segments, and what is the total length of roads running through protection zones? It is important to note that there may be a number of special cases. A road segment running along part of the boundary of a protection zone leads to a discussion of whether it should be counted. Even more special is the case where a road touches the boundary of a zone at only one point (e. g., a road junction). Such situations must be reckoned with and dealt with in a consistent manner. Also, the accuracy of the computational geometry is of interest. Taking into account that the geometry of geospatial data depends to a large degree on the accuracy achievable during data capture (Sect. 6.2.2), and that geospatial data that are submitted to a geometrical analysis may originate from different sources with different accuracies, we are bound to run into a few geometrically unfavorable situations where imprecision and sometimes even ambiguities result. Consider a road that touches a protection zone at one point. Due to imprecision, the system may arrive at an answer that there are two intersection points (the road is erroneously believed to pass through the zone, instead of only touching it) or no intersection points at all (the road is erroneously believed to remain outside the zone). Such inconsistencies may in most cases be overcome by setting fuzzy *tolerances*. This means that point locations that are closer to each other than a set value are considered to be identical. In this case, we can, for example, compute an average or weighted average of the coordinate pairs. Note, however, that the problems discussed above are inherent to information technology and GIS technology and cannot be totally eliminated due to the fact that imprecision is part of the game. Each data capture technology is limited in terms of the accuracy it can achieve, and the cost of data capture would explode if we tried to push the limit of achievable accuracy higher and higher.

At this point, it is necessary to delve a little into performance issues. In a polygon overlay scenario where a few polygons from one dataset are compared with a few polygons from another dataset, the results will be available in almost no time at all. However, the volume of GIS data is huge and the tendency for further increases is strong. Therefore, the time needed to compare thousands of polygons—each with potentially thousands of edge points—will become noticeable, up to the point where it interrupts the workflow and the user's line of thought in a very annoying way. So, we have to think about means to alleviate the effects of large data volumes. In the case of polygon overlays, this can be done by applying a special kind of spatial filter, or rather a special combination of such filters, to the datasets involved. A filter—we can also call it a sieve—can be seen as a device that retains a few gems and lets go of a lot of unwanted junk, as if one were looking for gold nuggets. The simpler the filter and the more unwanted things it can get rid of, the more suitable it will be. Spatial operations in general, and geometric analysis tools specifically, will gain considerably in performance if several such filters are applied in an appropriate sequence (Fig. 6.20). The sequence of a coarse filter followed by a fine filter is a good example. A coarse filter would be a simple approximation of a feature or a collection of features, which in most cases is a rectangle ranging from the lowest to the highest coordinate appearing in the feature or the collection in both the east and the north directions. This is called the minimum bounding rectangle (MBR). When comparing two datasets, as in the polygon overlay example given above, we could first compare the MBRs for each of the datasets and intersect them, arriving at a (smaller) rectangle. The filter will discard all elements that are outside this rectangle, while the elements in the intersection will need

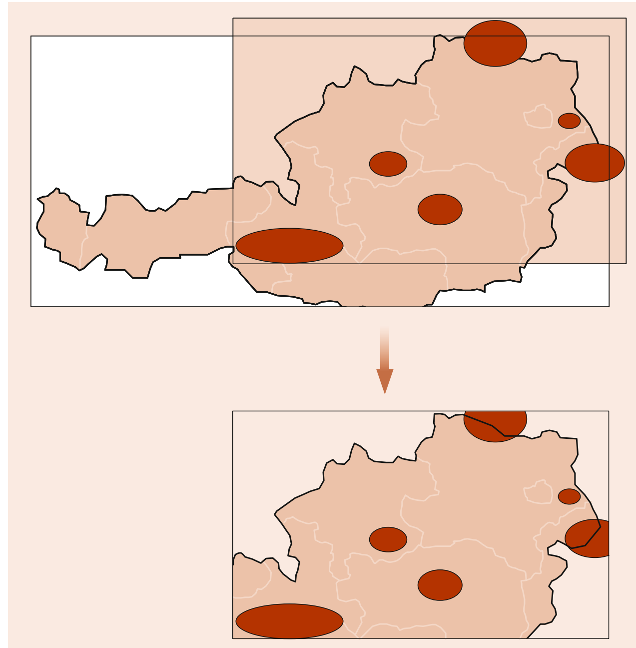


Fig. 6.20 Performance and recursive spatial filters

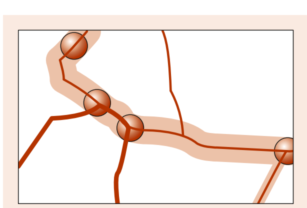
Fig. 6.21 A range query retrieves a false hit that will be discarded in a subsequent detailed point-in-polygon test

further scrutiny using a second filter. This second filter may be more complicated to apply, but it only has to deal with a small fraction of the original data. Such a strategy may be applied recursively. In this way, performance can be significantly improved because the rectangles are much easier to compare than the original geometries and because a lot of unwanted elements can be eliminated by this simple method.

A byproduct of these findings on the value of spatial filters is the concept of a range query. In GIS applications, we often need to compare a dataset against a rectangle. As an example, any display on a screen and any selection of geospatial data that can be drawn on a paper map will make use of a range query, since both are rectangles. Apart from this, range queries serve as quick and convenient tools to obtain preliminary results, and not only in geometric analysis. Note, however, that a range query may contain results that are discarded in a second, more detailed geometric query. These are known as false hits. In a query for cities in Austria, the initial range query will erroneously retrieve the city of Munich among many correct answers, such as Vienna. Munich must then be discarded in a second, more detailed geometric analysis: a *point-in-polygon* test (Fig. 6.21).

Among this array of geometric analysis tools, yet another type of spatial query, a *buffer analysis* (Fig. 6.22), is used in a score of GIS applications. In contrast to the methods for setting a comparison filter described earlier, buffer analysis typically involves using a line feature as an axis for a buffer, the width of which may be specified. Let us consider a line feature representing a road. We want to define a buffer that extends for a given distance on both sides of the road axis. We may want to find all agricultural areas that are within 500 m of the road axis or cities that are more than 10 km from a railroad line. Of course, a buffer around an area feature can also be defined, or one around a point (which is much simpler). In the latter case, it will be a circle. All these variations on buffering may be simple queries or they may be part of an intersection overlay performed to reduce the original dataset to features wholly or partially inside the buffer zone.

Fig. 6.22 Buffer analysis



As a final remark on these geometric analysis tools, let us point out the possibility of solving any of the problems listed above by temporarily moving to another world—the raster world. Taking the first example of overlaying two sets of polygons, we can perform a transformation from vector to raster for both sets, i. e., the administrative areas and the environmental protection areas (Sect. 6.2.5). For example, each administrative cell receives a number signifying the administrative unit to which it belongs, and likewise for the environmental cells. Depending on the accuracy to be achieved, the spatial resolution of these raster cells, which are needed only temporarily, may be rather coarse. However, it is essential that both datasets result in the same resolution, as they have the same origin. Now, all of the individual cells from both datasets can be compared in a cell-by-cell fashion. This results in a large number of comparisons; however, they are extremely simple to perform. If necessary, the results can then be transformed back to the original vector world, although in many cases the answers that were sought can also be obtained by counting and summing the findings from all individual cells. Raster technology methods can also be utilized in many other areas of geometrical analysis. A buffer can, for example, be described by first transforming data into raster format and then inflating the central axis pixels by a given amount.

To sum up, geometrical analysis functions can be characterized in the following way.

- They are at the core of many analysis processes, since geometry is a basic aspect of geospatial data
- Polygon overlay, intersection, range queries, buffering, and nearest-neighbor searches are examples
- Spatial filters in special arrangements help to boost the performance of geometric operations.

Topological Analysis

Topology is a branch of mathematics which deals with spatial properties and relationships that remain intact during continuous deformations such as stretching. A balloon metaphor is often used. Imagine a balloon that has a map of Europe drawn on it, or a road network. If you inflate the balloon, the geometry changes but adjacent countries remain adjacent, the tiny state of San Marino will still be inside Italy, and roads connected to each other via a junction will remain connected. Also, a cube can be continuously deformed into a sphere, but never a donut. Cubes and spheres are topologically equiva-

Fig. 6.23 Nodes and edges of a street graph in a routing application

lent to each other, but not to donuts. Topological properties such as adjacency, containment, and connectivity, though often implicitly known to the user, are not explicitly modeled from the beginning. Instead, topology must be built through semiautomatic processes such as *clean and build* during the data capture or update phase. From this point on, topological properties should be maintained and supervised in a GIS, since they are not only an essential part of the conceptual data model but also a prerequisite for many topological analysis tools, some of which shall be explained here. For further reading on this topic, please refer to Chap. 12.

Routing is the process of selecting paths in a network along which to send network traffic. This general definition encompasses several types of networks such as powerlines or telecommunication landlines or wired networks carrying electronic data. In the GIS domain, it is mainly transportation networks such as roads that are subjected to routing analysis, although utility networks would be applicable too, since they have many similar aspects to transportation networks, such as capacity. Routing is heavily reliant on having the correct topology of the road network. *Topological nodes* correspond to road intersections or junctions; *topological edges* lead from one node to another, so they correspond to road or street segments between two intersections (Fig. 6.23). Topological consistency means that there is no ambiguity between any two nodes or any two edges in the network and that each edge has a node at each of its ends, while any node may have several edges starting and/or ending there. In special cases, there may be only one edge connected to that node (the end of a dead-end street). In rare cases, there may be isolated nodes (with no edges at all) or nodes supporting exactly two edges (these should not really be nodes). A road network should also be connected. This means that, for any two nodes in the network, there should be at least one *path* (an ordered sequence of edges) connecting them. In general, there will be more than one path. Another term used for these networks is *graph*. Topology is therefore also dealt with in graph theory [13].

A special case is a network where, for any two given nodes, there is only one path connecting them. This is called a *tree structure*. For a road network, this would be an odd structure, but (for example) river networks are always represented by tree structures. The topology of a road graph is initially independent of geometry. Coordinates enter the picture only at a later stage (if at all), as well as lengths and distances. An example where such lengths and distances are mostly irrelevant to the user is given by subway line maps. It does not matter how long the tracks are; only whether and how two stations are connected and at which station a change of trains is possible and/or necessary. This is pure topology. In contrast, any road traffic routing application on the Internet has already taken one more step towards introducing some geometry into the picture—supplying each edge with an indication of its length (which may be the mileage or time required, or some other measure). Starting out at a given node with the aim of reaching another given node, the shortest path can be found in the simplest (yet least effective) way if we use the picture of a flat surface that has all roads engraved like canals onto it, and then you pour water onto the starting node. Assuming no loss due to friction and seepage, the drops of water will move through all the canals at constant velocity. The first drop of water to reach the end node has taken the shortest route. Algorithms that were designed to achieve greater efficiency (for example, the famous *algorithm of Dijkstra* [14]) will find their way through a network of edges and nodes much faster than this. Another increase in speed is achieved by searching simultaneously from both ends. Then there are scores of heuristic (not theoretically correct but practically successful) methods that, for example, restrict the number of nodes and edges to be visited by applying heuristic rules of thumb. So, when going from Paris to Rome, it is highly unlikely that edges to the northwest of Paris or to the southeast of Rome have to be taken into consideration. Other refinements to routing may consider the road hierarchy and therefore also the path-finding hierarchy. Taking a highway is usually better than taking provincial roads, even if the resulting geometric length may be significantly greater. This leads us into a discussion of *shortest* and *best*. In many cases, the shortest way may not be the best way. Routing algorithms take into account many aspects, putting different weights on decisions such as whether you prefer autoroutes or more scenic routes, whether you want to have a quiet (but long) or a rather hectic (but short) trip, whether you are an average driver in terms of speed, etc. Taking the means of transportation into account (pedestrian, bicycle, private car, public bus, taxi, or emergency vehicle) may also improve the suitability of the chosen routing model. Finally, complex street intersections can be modeled in a refined way, taking into account any restricted or recommended maneuvers, such as left turns, right turns, U-turns, etc.

Routing algorithms form the basis for several other important topological analysis tools. *Accessibility analysis*

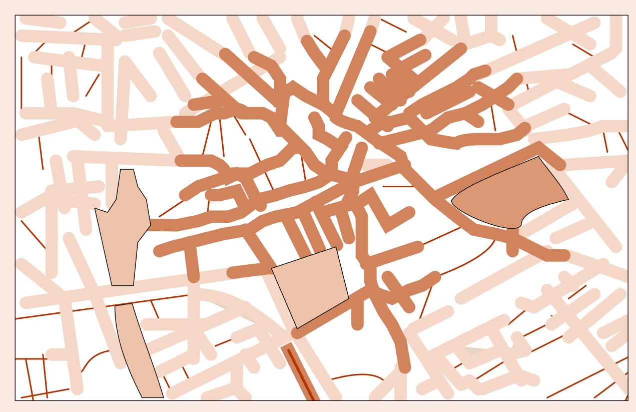


Fig. 6.24 Accessibility analysis: all street segments within a given distance of a center

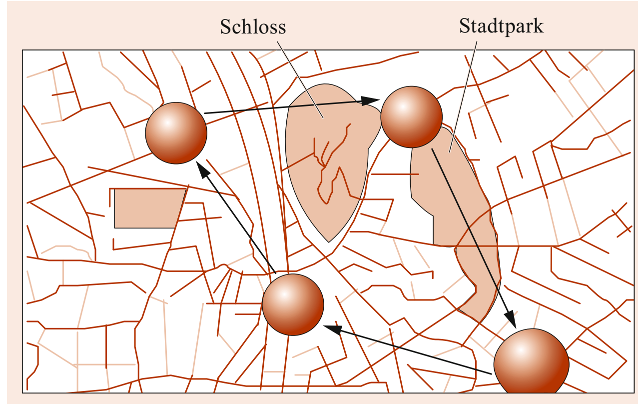


Fig. 6.25 Traveling salesman problem: finding the least costly route when visit a given number of targets

6

(Fig. 6.24) gives all of the targets that can be reached in a given network within a certain distance or time (or some other metric) of a given home base (e. g., all restaurants within a 5 min bicycle ride, all ski trails with an entry point and parking lot within 2 h of driving, or all stops of the public bus service that are within 300 m walking distance). Clearly, these are reality-checked, down-to-earth alternatives to a Euclidian metric that would use circles around the home base, modeling distances as the crow flies. If we reverse the roles to perform an accessibility study that asks, for example, which (and how many) households in a city are within 300 m of a public bus stop, where there are bottlenecks in this service, or where there is an unnecessary excess of bus stops, the same methods can be applied; only the interpretation of the results has to be adapted.

The *traveling salesman problem* (Fig. 6.25) is another prominent example of a topological analysis built upon routing. Consider a salesman who must visit several spatially dispersed clients and must find the optimum path with the lowest costs, where the *costs* do not have to be monetary. Of course, traveling fewer kilometers, spending fewer hours traveling, and paying fewer toll fees will reduce the costs in their strict sense, but a lot of other parameters, such as the required resting time, whether there are a sufficient number of hotels available at good prices along the proposed route, whether the route is scenic, and a desire to minimize any additional inconveniences, can influence the definition of costs. In fact, this will influence the weighting of the edges of the graph. Traveling salesman problems can be arbitrarily requested to render a closed-loop solution. They can furthermore be differentiated into a type where the order of client visits is fixed at the outset and another type where this sequence can be determined by the system. Of course, the first kind is simpler, being nothing more than a series of routing problems. The second type is much more ambitious. If there are more than, say, ten targets, such a problem can only be solved by heuristic methods, since the number of mathematically possible combinations simply explodes. Applications of this sort also include a postman delivering letters to the postboxes in his area (even though this is not a problem they will need to solve anew every day), a taxi delivering pizza to several customers in a row, a parcel delivery service, and a municipal garbage collection service.

Let us conclude this section on topological analysis functions with the scenario of *fleet management* applications (Fig. 6.26). If individual postmen delivering letters are collectively considered to be an armada of postmen; if individual garbage trucks are assembled into a municipal fleet of garbage collection trucks; or if there are a fleet of taxis that are requested using the same phone number, we arrive at fleet management. In this case, a theoretically strict and sound solution is impossible to achieve due to the huge diversity of combinations possible. Since, for each truck in the fleet, heuristic suppositions have to be made regarding the sequence of individual households to be visited, the management of a whole fleet of such trucks will be an even more complicated matter. Checking the impacts of various changes to the original setup is where GIS technology can

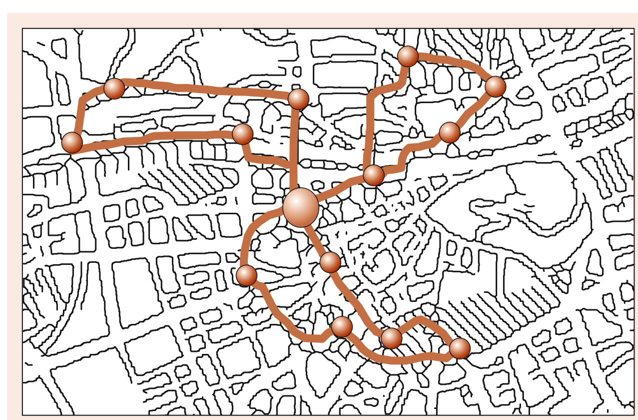


Fig. 6.26 Fleet management: several vehicles on several routes are available

be utilized, leading to a solution where the strengths of GIS analysis tools (repeatability, precision in computing, consistency) can be combined with the strengths of the expert using the GIS: experience, competence, and a certain amount of intuition.

To sum up, topological analysis functions can be characterized in the following way.

- They presuppose a *topologically correct data structure* that is not inherent; it must be set up
- They most often also comprise *metric aspects such as distances*
- Routing, traveling salesman problems, accessibility, and fleet management are prominent examples.

Interpolation and Approximation

In the first part of this chapter (Sect. 6.2.2), we discussed the modeling process and the resulting model of geospatial data as a way of defining a partial replica of the real world that is an approximation to it. As we have frequently been using the example of a mountain road to clarify aspects of GIS technology, we will again use it here. The infinite number of aspects of a real-world road need to be reduced to a feature that carries a classification mark (for example *road*) and has a finite number of axis points (given by its coordinates) and a finite number of attributes (for example, the attribute *road name*, with the value *Semmering-Straße*). The question of which points should be chosen to represent the geometry in the best possible way remains. We cannot choose a large number of points, as this would contradict the overall modeling goals and would also unduly increase the data volume. Clearly, any mountain road will have some characteristic sections (those with serpentine curves), while many other sections are (almost) straight. It will be appropriate to choose more points at or near the curved parts, while the other parts of the road may only occasionally need to contribute a point. The resulting skeleton of control points can then be stored in the GIS database. If a GIS application subsequently needs to visualize the road, it can lay an *interpolating line* or *curve* along those control points (Fig. 6.27). This is usually done in 2-D when producing a map-like portrayal, but it can be extended into 3-D. Interpolating curves are provided by GIS systems on a standard basis, with parameters that can be set by the users. The simplest form of interpolation would be achieved by connecting the control points with straight lines. For the case of city streets, where the control points are often set at street junctions, this linear interpolation would even be the best solution. For a mountain road, a higher-degree interpolation may be preferable. Quadratic or cubic polynomials are often used as building blocks. Typically, an interpolated curve consists of a series of such polynomials, each defined in only a limited range, but certain smoothness conditions have to be met at the joints where they are welded together. Among the whole range of possible realizations, spline curves are the best-known representatives because they show favorable behavior with a minimum of undesirable oscillations.

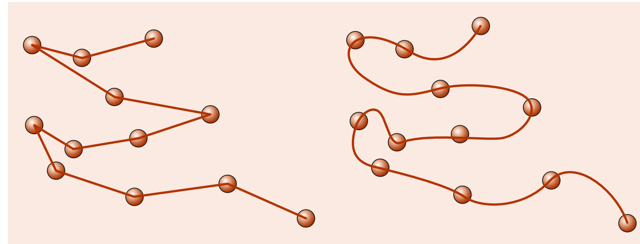
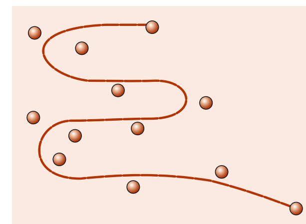


Fig. 6.27 Interpolating with straight lines (*left*) and curves (*right*)

Fig. 6.28 Approximating curve



If we modify the restriction that the resulting curve should go through the given control points by allowing for small deviations, we arrive at an *approximation* (Fig. 6.28). In this case, we have to define what we mean by a *small deviation*. Here, statistical methods enter the picture. In many cases, an approximation is deemed to be appropriate if the sum of the squared residual errors is a minimum. The aspects of such a concept are nicely illustrated if we assume the simple case that, for all control points in a given section of the mountain road, an approximating straight line should be found that is *best* with respect to the condition just mentioned. Note again that linear interpolation will result in a zigzag line passing exactly through all the points, while in the approximation case, the resulting line will pass *near* to all the points. In statistics, this method is called *linear regression*. The condition that the sum of the squared residual errors should be minimized gives rise to the term *least-squares regression*. Of course, this implies that there are also other forms of regression, depending on the conditions we want to set for the estimator. All this can be seen as a part of *adjustment theory*, with an apparatus of formulae that are well known and applicable in many other fields of GIS technology and adjacent geosciences. Another neighboring application would be the evaluation of *trend lines*, which makes use of the same methods.

Let us now move on to the interpolation and/or approximation of surfaces, which is, for example, necessary to produce a *digital terrain model* (DTM). In the previous paragraphs we described how given control points may lead to an interpolating or approximating curve. Now we just add another dimension, but the procedure remains the same. Let

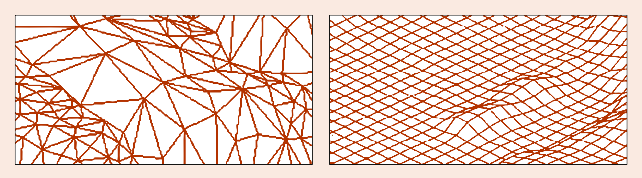


Fig. 6.29 Digital terrain models: TIN and grid

us first consider the interpolation case. Instead of connecting basic line segments to form a curve, we need to consider two-dimensional elementary patches. Triangles are favorable because they are planes, allowing for linear interpolation. A digital terrain model composed of triangles will look like the surface of a gem, with triangular facets (Fig. 6.29). However, the vertices of triangles are mostly distributed in an irregular fashion. This is why such a terrain model is called a triangulated irregular network (TIN). An alternative would be to use patches whose projections onto the plane given by the first two coordinates would be rectangles. The advantage of regularity is balanced by the drawback that the resulting quadrangle patches are no longer linear but at most bilinear, since four points usually do not constitute a plane. This alternative (Fig. 6.29) is called a grid DTM. Both alternatives appear in practical applications. Grids are the standard form for larger areas with little topographic diversity and moderate demands on the geometric accuracy of heights, while TINs are deployed for those portions of a terrain where break lines, edgy ridges, sharp peaks, etc., occur, and especially where the accuracy of geometric details is important.

Terrain models, like curves, may also move from being an interpolation to an approximation, where the resulting composite surface passes near to the given points but not exactly through them. The criterion for closeness may again be a least-squares minimum, so the apparatus mentioned above will also apply here. The terrain surface may technically be handled like a *trend surface*, using appropriate adjustment techniques.

Let us note that the approximation of curves and surfaces is not only a GIS topic. CAD and other desktop graphic applications make use of the same concept. Designers—whether designing car bodies or character fonts or pursuing other design goals—all have similar needs and therefore use similar methods.

Terrain models are a standard tool in GIS analysis since the geometries of many geospatial features are modeled in 2-D only. For example, cadastral features such as parcels only have coordinates in two dimensions, the third dimension being irrelevant. If the owner of the parcel digs into the ground, it is still the same parcel. If a utility line is planned that passes 5 m above the ground, it is still the same parcel, with all the legal conflicts that may originate in this example. For many GIS applications, the parcel has to be *lifted* from the base plane to the current terrain so that the parcel area seems to be coating the terrain. Note that this formulation uses a visualizing paradigm, but of course the problem is not restricted to visualization functions. So, in countless situations, for a given (x, y) pair, a z value has to be estimated. In both the TIN and the grid cases, this estimation can be done by first finding the appropriate patch (triangle or quadrangle) and then interpolating in a linear (or bilinear) way. Of course, such a strategy can only work if the terrain model is available. Some preprocessing has to be done to set up an appropriate DTM. Also, every new terrain point that enters the GIS will, in principle, call for a DTM update. For data such as those in a GIS, which have a tendency to change frequently, this may be cumbersome. In such cases, alternatives are preferable, such as the *moving average* approach. Here, for each point whose z coordinate needs to be estimated, a bunch of available z values in the neighborhood are selected on the fly and an average value is computed. Clearly, the size of the neighborhood will influence the degree of compliance of the new point height with the overall terrain. If, in an extreme case, all points of the whole dataset are taken, all new point heights will be the same. This would be a global effect. The smaller the neighborhood, the more local the effect. Often, the simple average is replaced by a weighted average. Some of the neighboring points can make larger contributions than others. The weighting is, for example, done by using the inverse of the distance to the new point as a weight. Points far away then have less influence than nearby points.

Let us conclude this section on analysis functions by naming a few models deriving from DTM. *Contour lines* (Fig. 6.30) are a well-known graphical element of topographic maps. They connect all the points on a map at the same altitude. Contour lines are drawn for specific altitudes (e. g., 700 m, 750 m, 800 m, etc.) separated by a given altitude interval. Such lines can also be extracted from a DTM. The procedure is simple but lengthy. Let us assume that a TIN is available. For each TIN triangle, the heights of the three vertices are compared with the altitude of the contour line to be extracted. If the line is below or above all three vertex heights, the triangle can be neglected. Otherwise, for this altitude, points of intersection along the triangle's sides have

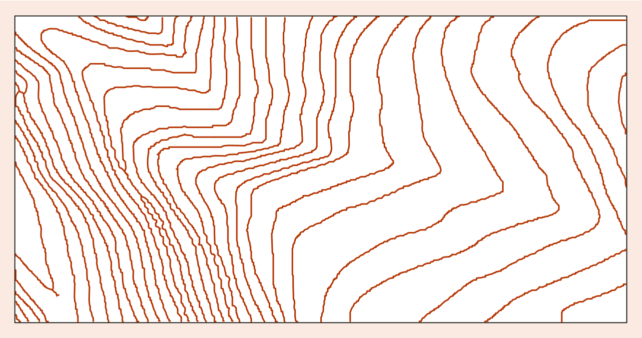


Fig. 6.30 Digital terrain models: contour lines

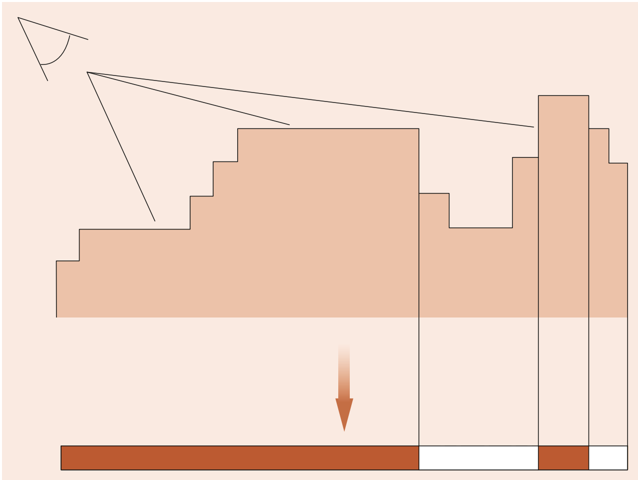


Fig. 6.31 Viewshed analysis of a terrain model to identify its visible portions

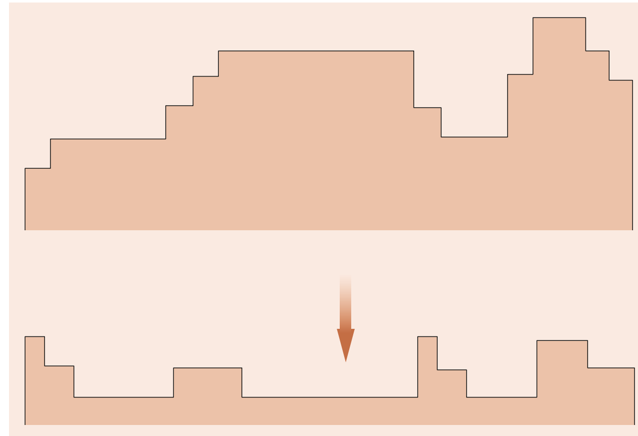


Fig. 6.32 A terrain model and the corresponding slope model in profile view

to be computed. Naturally, there can be at most one such intersection point per triangle side. Also, since contour lines cannot disappear or appear suddenly, some other restrictions can be applied. In the end, when all the intersection points have been computed, they have to be arranged as a *string of pearls*, again observing some natural rules.

Viewshed analysis (Fig. 6.31) is another tool based on the existence of a DTM. Given a viewing point, we may ask which portions of the terrain are visible and which are hidden behind hills in the foreground. The method for finding the answer is conceptually easy. From the given viewpoint, a line of sight needs to be drawn to every patch of the DTM, be it a triangle or a quadrangle. This patch is visible if the line does not encounter any other patch before arriving at the patch in question. Viewshed analysis can also be used as an instrument in several other ways; for example, when choosing a viewpoint that corresponds to the position of the sun. In this way we can find out whether a given patch of the DTM is currently sunny or shady, leading to a complex analysis of how much sunshine and how much shadow should be expected in the course of a year. Also, the line of sight can be reversed. The viewpoint takes on the role of a television or cellphone transmitter, and the question is then: from which patches of the DTM is the transmitter visible? There are also applications that aim to find locations for smoke stacks that will minimize their visual impact on a tourist village.

Another offshoot of the DTM is a *slope model* (Fig. 6.32). Determining the change in height in each part of the original DTM is just a natural extension of the high-school math that we were taught to find the derivative of a function. Where the function is steepest, the derivative is highest; where the function has a local maximum or minimum, the derivative is zero. Here, we just generalize this to a situation where there is a function of two variables such as $z = f(x, y)$. This method is conceptually easy to grasp.

Slope models have some favorable aspects. When visualizing them, they provide a good perception of steepness, even though it is not readily apparent whether the slope goes up or down. Also, the perpendiculars to contour lines are the lines of steepest descent or ascent. This, in turn, makes it possible to search for the places where adjacent perpendiculars join. Runoff analysis based on precipitation models makes use of this. Consider drops of water and how they behave in an idealized environment in the absence of seeping or evaporation. They run down along lines of steepest descent, eventually meeting other water drops, so we can predict the paths of the small streams of water, creeks, and rivers that form ad hoc during heavy rainfall.

An *aspect model* will additionally show slope orientations (i. e., whether they are oriented north, east, west, or south). The resulting slope or aspect model is identical in structure to the original DTM. This leads us to the conclusion that the DTM and similar models (such as population density models used in geomarketing, temperature models, and precipitation models) can be subjected to a whole range of arithmetical and/or logical operations, again leading to yet another model that is structurally equal to a DTM. Consider, as an example, a DTM for a waste disposal pit on a day in the year 2018, and another DTM for the same pit 1 year later. Computing a *difference model* (Fig. 6.33) by simply subtracting the values in corre-

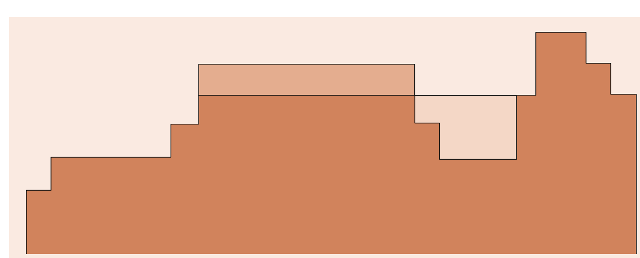


Fig. 6.33 Two terrain models and their difference model in profile view

sponding patches will lead to an assessment of the increase. In the same way, we can compute sums or logical comparisons. Another example would be the combination of terrain, aspect, and population density, all modeled by DTM technology. A southwestern slope at medium elevation with few or no people around would be an ideal spot for a holiday cabin.

To sum up, analysis functions for interpolation and approximation can be characterized in the following way:

- Interpolations and approximations for lines and surfaces (DTMs) are at the core of many other GIS analysis functions
- They give reasonable estimates for heights based on interpolation and approximation assumptions
- DTMs can be used to generate contour lines, slope and aspect models, viewshed models, and others.

Planning and Simulation

A GIS lends itself perfectly to planning and simulation. The *what if* question expresses a deep-rooted need that mankind has to predict the future and to determine which parts of the future can be influenced to a certain extent. It is therefore natural that tools such as GISs should be used for such a purpose. Since, in any planning process, the ratio of what should be created anew versus the number of preexisting features will be quite small, it is possible to define several scenarios for possible realizations of a given project. The additional effort required to define any further scenario is quite small, since most of the geospatial data and functions needed for this extra alternative will already be available. Urban planning and infrastructure projects such as new roads and railway connections first come to mind when one thinks about planning and simulation. Several different versions of a new housing project, a new shopping mall, a new land improvement zone, as well as new roads, highway exits, city tunnels, etc., may be created and presented to the public for discussion, allowing for more direct democracy and the contribution of new ideas that allow further improvement. Of course, this has always been done by architects and urban planners, who provide plans and graphical views of the planned scenarios as the basis for such discussion. GIS technology not only makes it easier to increase the range of scenarios considered; it also offers new ways of visualizing them, such as a virtual 3-D city walkthrough that leads to a more realistic evaluation of the different alternatives, as it highlights their visual impact and how well they fit into the neighborhood. When mixing reality and virtual reality, we arrive at *augmented reality* (AR), which is especially promising during planning processes. As an example, placing a virtual, as-yet nonexistent structure into an existing landscape can significantly aid the decision process.

Additionally, GIS analysis functions can be used to evaluate the consequences of different alternatives in depth. Consider, for example, a new shopping mall to be created in a suburban area. The costs of land, construction, and infrastructure supply (roads, electricity, water, gas, telecommunication, and sewage) can be estimated for each alternative, and the compliance with regulations such as those for zoning, construction, and environment can be assessed, provided that the corresponding information has been modeled in terms of geospatial data. However, GIS analysis tools can provide much more information that is important in the decision process. Based on an estimate of the expected traffic frequency, the impact on access roads can be modeled. Their capacities can be tested and the overall traffic increase can be predicted. Environmental issues such as pollution, noise, and contamination also enter the picture. The overall cost of such a project alternative—not only in terms of cash, but the total burden, ranging from commercial to environmental aspects—may be assessed in comparison with the benefits. All of this is of great help during the process of decision-making. A GIS is sometimes defined as a tool for decision support, and the applications described here strongly emphasize this definition of a GIS.

The question of the location of a new shopping mall can be seen as a special case of the *optimum placement problem*. This is a rather general term. For example, it also covers the distribution of public bus stops in a city, as was discussed earlier in the context of topological analysis functions. Bus stops, pedestrian crossings, fire hydrants, train stations, and autoroute exits are just a few special points of service for which the optimum location may need to be considered. Another variant of optimum placement that does not depend on linear networks of roads and infrastructure utility lines is the selection of the optimum locations of transmitters for television or cellphone network providers to achieve continuous coverage. Since line of sight is relevant in these cases, the solution to these placement problems depends to a large degree on the terrain and also the manmade structures (buildings) on the terrain. A digital terrain model (DTM, see the previous section) and an appropriate situation model for the buildings will therefore be necessary. However, we also need to consider the distribution of the potential recipients of the television or phone signals. The number of households per administrative unit or even per city block can be provided by statistical bureaus operating at the state or federal level. Often they are also commercially available. The household density per given square cell (for example, within a km^2) can be interpolated in the same way that a terrain model is created, as described previously. We arrive at a population density model that looks very similar to a terrain model. The peaks in the population density model denote heavily populated areas; valleys and lowlands indicate sparsely settled areas. Comparing the models for the terrain (plus buildings) and the population, we can begin by placing the transmitter onto each of the terrain peaks and computing how many

Fig. 6.34 Environmental disaster mitigation. Floods and emergency plans for the evacuation of areas according to assigned priorities. Numbers *1*, *2*, *3* mark medical centers and their surrounding areas with an according evacuation priority. (Source of base data: City of Graz Geoportal. Overlay: Institute of Geodesy, TU Graz)

households in *population peaks* can be reached by direct line of sight from each terrain peak. Of course, the eventual solution will also depend on other factors—environmental, zoning, accessibility by road, visual impact, etc.—that can also be modeled and evaluated by GIS analysis tools.

Having already explained how a population density terrain can be modeled and put to use, we now consider a further field of simulation and planning applications. The term *geomarketing* is used to address issues such as the population analysis of factors including age, income, education, cultural background, mobility, and many others that are covered by statistical bureaus and agencies; but also markets and consumers, target groups for special products, the placement of stores, shops, and malls, the presence of competitors, etc. Many of these applications can be modeled in the same way as for the household density in the previous example. One of the principal geomarketing paradigms is *tell me where you live and I'll tell you what you buy*. There is significant coherence in the consumer behavior of residents of neighborhoods in city centers. This is also true of residents of suburbs, and of rural residents. This fact is exploited and put to use by geomarketing applications.

Another wide range of simulation and planning applications that can be efficiently handled by appropriate GIS analysis tools are *environmental disaster mitigation* plans and the corresponding *emergency and rescue* scenarios. As an example, for rivers and streams in an area of interest, hydrological models can be implemented that will supply an estimate of the extent of the inundated area and the associated depth of water when the water influx exceeds a certain level. Typically, there will be statements about high water levels occurring on average every 30 years, 100 years, etc. A hydrological model needs as input a lot of geospatial data such as terrain models and the locations of buildings, roads, bridges, dams, and sealed areas. This is because hydrological modeling is a complex issue, and it cannot be expected that the water will rise gradually, like in a bathtub. Instead, the one-, two-, and three-dimensional dynamics have to be taken into consideration, as well as the influence of obstructive objects such as dams or bridges and the ability of the soil to soak up water. The results of the hydrological modeling are directed back into the GIS, leading to further analysis. Additionally, meteorological models can be introduced, as well as observations of flood levels done on the fly by emergency and rescue squads via location-based service (LBS) devices can be used to calibrate the initial model and for model iteration, leading to better and more realistic results (Fig. 6.34).

Computations using such a complex model will involve huge amounts of data. They are extremely time consuming. A real-time environment is therefore not feasible. Instead, most of the computing must be done beforehand, and the results stored for retrieval when the need arises. This can be done by classifying typical situations and, for each of these situations, having a model and its solution to hand. Knowing the water rise that will occur under certain preconditions, we can arrange for measures such as rescue plans to be triggered in the event that those preconditions are fulfilled. GIS analysis tools are highly suited to this task. Consider, for example, that the expected water level for an inundated area is known to an accuracy of 20 cm. If geospatial data within that height accuracy range are available for buildings, streets, and bridges in that area, we can decide which of the buildings are still accessible to certain types of rescue vehicles, and—using routing functions, as previously described—we can determine which routes will still be accessible. By classifying buildings based on their priority (schools, hospitals, and nursing homes will have the highest priorities), we can further enhance the realism of this analysis tool.

As we approach the end of this section dealing with simulation and planning, we arrive at *agriculture and forestry*. Having concentrated on floods in the previous paragraph, we now focus on the opposite—droughts, which pose a significant threat in an increasing number of areas around the world. A statistical increase in periods of extreme weather conditions can be observed even in zones of moderate climate. The effects are influenced by several parameters, such as the terrain, its height, its steepness, and its aspect (a north-, south-, east-, or west-facing slope), the average temperature, evaporation, runoff, and soil, and a number of other factors. Readers may ask themselves how much simulation and planning we can introduce into an application where most of the parameters are fixed or outside our sphere of influence. However, there is a certain amount of choice with respect to the crops to be cultivated, the number of harvests per year and their timing, and the fertilizers and other chemicals used. This justifies the inclusion of such applications in the discussion in this section.

To sum up, analysis functions for simulation and planning can be characterized in the following way:

- They provide an efficient means to model different variants of a planned project
- Visualizing them brings interested parties closer to the decision-making process
- New technologies such as AR (augmented reality) enhance the meaningfulness of alternatives
- By also analyzing secondary effects, we become aware of future impacts of current activities.

6.2.9 Design and Presentation Functions

The last group in this list of GIS functions consists of functions that are typically useful when data have passed through all of the other stages in their lifecycle, including capture, structuring, transformation, storage, and analysis. Presenting analytical results graphically is a highly efficient way to express spatial interactions, patterns, and correlations. Due to synoptic effects in visualization, spatial knowledge can be gathered and transmitted in a straightforward manner. In some cases, the visualization of analytical results may mark the end of the GIS specialist's work. The commissioned work is then handed over to the patrons that have requested it; the specialist relinquishes control and collects the financial compensation. As GIS functional tools become increasingly sophisticated, the paradigms change, and visualization becomes the core of the analysis process. This allows for iterative refinements and changes to the overall model for the solution strategy and to parameters and constraints. Especially in planning scenarios, trial and error is becoming increasingly important. Sand-table exercises considering the question of *what if* are inexpensive techniques for examining the consequences of projects before any money or other effort has been put into them. However, there are also other scenarios where playing with visualizations is useful. Consider, for example, the application of image-processing software to raster images, as we might do to photos from our last vacation trip—changing the illumination or spreading the spectrum of gray values may bring out details that were hidden before. In a GIS environment, features and patterns may become visible in this way. Ideas such as this eventually lead to augmented reality, where, in this playful mode, we may decide to add to those aspects we already see (such as an orthophoto) an overlay of administrative boundaries and/or subterranean utility lines, etc.

At this point, it is necessary to address a fundamental difference between geospatial data and visualization that is quite often neglected when dealing with geographic information. Geospatial data are by their nature nongraphical; for example, geospatial data describe a road through its geometry (the point geometry of its axis vertices, and perhaps also the road width at this point and the height above sea level) and its semantic attributes (administrative relevance, suitability for certain vehicle types, paving, etc.). In a graphical visualization, the geometry is mapped to a polygon or interpolated curve and the semantic attributes are mapped to graphical parameters such as line width or the appearance of two lines running in parallel at a given separation. These are mapped to colors, with each color denoting a different administrative relevance, such as a state road, a provincial road, or a local road. They are also mapped to different symbols accompanying the visualized line, such as a crossed-out camper symbol conveying the information that this road should not be used by campers. Visualization is often accompanied by cartographic generalization, since only by omission, simplification, combination, exaggeration, and displacement (the principal generalization tools) can a satisfactory appearance be achieved, such as one that avoids cluttering and interference from cartographic symbols to increase readability and reduce information loss.

Humans usually come to understand the meaning of visualized geospatial data through learning and experience. This may lead to the problem that the visualization of data is taken to be the data itself. A typical consequence is a digitization of a paper map that has undergone a cartographic generalization process and later serves as the basis for a digital dataset. This dataset contains geometric errors that were introduced by the generalization.

In addition, map-reading skills and experience may not be the same in all parts of the world, which can lead to different interpretations of the map content.

This is why, in principle, there should be *two worlds* that are separate yet linked: the world of geospatial data, which resides and is analyzed in a GIS, and the world of data visualization. The term *portrayal* is used for the second world (Chap. 13). Many national topographic agencies are using this twin strategy. Preferably, for each geospatial application database, there should be at least one portrayal, and sometimes more than one, because changes in scale, purpose, and media (paper or screen) call for alternatives. Geospatial data are sometimes said to be scale independent. This is not really true. Data describing reality at the national level (for example, in terms of roads) are quite different in structure and semantics than data describing *the same* reality at the local level, for example in a city. Therefore, scale independence only holds for a limited range of adjacent scale factors. Portrayal data are still far more scale dependent than geospatial data because minor changes in scale can result in strong graphical conflicts. For coherence and update purposes, it would be desirable to have direct links between corresponding data, but in practice this is rarely the case. If visualization is seen as an end product in the lifecycle of geospatial data, this is not such an important flaw. However, if the visualized world is used as a sandbox playground (as discussed above), which leads to better solutions by allowing the interactive manipulation of features, it is of course necessary to have a direct link back to the geospatial data or, alternatively, allow for another compromise and reduce some of the requirements we just used as arguments in favor of cartographic processing.

Let us focus on another question that directly addresses the overlap and conflict between the two worlds mentioned above: the question of visualized texts. Strictly speaking, any visualized text belongs exclusively to the second world of portrayal. Its counterpart in the first world, the geospatial world, is, in most cases, an attribute value. Consider, for example, a database of municipalities in a certain country. The municipality class has an attribute called *name*. Instances of this attribute include *Alpha City*, *Betatown*, etc., and these instances will be visualized in the portrayal, with the text displayed near to the reference point of a municipality using a text font reserved for municipalities and at a text size that corresponds to the population size, which is another attribute value of the municipality class. Also, in principle, any GIS application could refrain from showing text, since we can always find the value of any attribute (in this case the name of the municipality under the mouse pointer) with a *mouse-over* effect. However, cartographic texts are something that we become accustomed to after we first learn to interpret maps in school. This is why cartographic texts are sometimes treated as not only aspects of the portrayal but also a special kind of geospatial feature.

This leads us to a final discussion point here. The question arises of how much we should cling to old traditional ways of displaying geospatial data. The most appropriate example is again the topographic map, which we become accustomed to during our lives. If we compare the traditional graphical appearance and the traditional legends for symbols, line types, hatch styles, etc., with the methods used in of present-day computer graphics, including animation, pseudo 3-D, and the introduction of sound, it may seem advantageous to introduce some of these new technologies so that visualizations of geospatial data can benefit from them. However, occasionally there are standards and rules for such symbol catalogs that cannot be neglected. An even bigger argument is the fact that human experience, human habits, and expectations will never change as fast as technology. The overall goal of visualization will always be to minimize the loss of information as data are transported from the person conceiving an application and its visualized form to the person receiving it, and retrieving as much information from it as possible. Sometimes adhering to old but approved and familiar ways may be preferable. More on issues relating to visualization and cartography can be found in Chap. 13.

To sum up, design and presentation functions can be characterized in the following way:

- They deal with portrayal data differently from their counterpart, geospatial data
- These two worlds are preferably linked
- Portrayal is often the end product in the lifecycle of geospatial information
- However, visualization increasingly extends into the core of GIS analysis, offering a sandbox scenario.

References

1. Bartelme, N.: GIS-Technologie: Geoinformationssysteme, Landinformationssysteme und ihre Grundlagen. Springer, Berlin, Heidelberg (1989). in German
2. Bartelme, N.: Geoinformatik: Modelle, Strukturen, Funktionen. Springer, Berlin, Heidelberg (2005). in German
3. Dueker, K.J.: Land resource information systems: a review of fifteen years of experience. Geo Process. **1**, 105–128 (1979)
4. Burrough, P.A.: Principles of Geographical Information Systems for Land Resources Assessment. Clarendon, Wotton-under-Edge (1979)
5. Göpfert, W.: Raumbezogene Informationssysteme. Wichmann, Heidelberg (1991). in German
6. Cowen, D.J.: GIS versus CAD versus DBMS: what are the differences? Photogramm. Eng. Remote. Sens. **54**, 1551–1555 (1988)
7. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W.: Geographic Information Science and Systems, 4th edn. Wiley, New York (2015)
8. Zehnder, C.A.: Informationssysteme und Datenbanken. Teubner, Wiesbaden (1989). in German
9. Goodchild, M.F.: Citizens as sensors: the world of volunteered geography. GeoJournal **69**(4), 211–221 (2007)
10. Pavlidis, T.: Graphics and Image Processing. Springer, Berlin, Heidelberg (1982)
11. Pratt, W.K.: Introduction to Digital Image Processing. Taylor + Francis, New York (2014)
12. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 4th edn. Pearson, New York (2018)
13. Bondy, A., Murty, U.S.R.: Graph Theory. Springer, London (2011)
14. Dijkstra, E.W.: A note on two problems in connexion with graphs. Num. Math. **1**, 269–271 (1959)

Norbert Bartelme Professor Bartelme, before his retirement, was Head of the Department of Geoinformation and Dean of Studies for Geomatics and Geospatial Technologies at Graz University of Technology, Austria. His current research fields are geodata standardization and interchange, spatial information systems, interoperability in the field of geodata and geographic information systems (GIS), XML-technologies for the use in location-based services, and GIS risk management.

Change Detection

7

Jérôme Théau

Contents

Abstract

This chapter is an overview of change detection techniques applied to Earth observation. Section 7.1 gives a short definition of change detection followed by the historical background of this process (Sect. 7.2). Methods of change detection are described in Sect. 7.3, and are illustrated using an analysis of wildlife habitat (Sect. 7.4). Typical applications are then listed. The chapter concludes with a discussion of future directions of change detection studies (Sect. 7.5).

J. Théau (✉)
Université de Sherbrooke
Sherbrooke, QC, Canada
e-mail: Jerome.theau@usherbrooke.ca

7.1 Definition

Change detection can be defined as the process of identifying differences in the state of an object or phenomenon by observing it at different times [1]. This process is usually applied to Earth surface changes at two or more times. The primary source of data is geographic and is usually in raster format (e. g., satellite imagery), analog format (e. g., older aerial photos), or vector format (e. g., feature maps). Ancillary data (e. g., historical or economic data) can also be used.

7.2 Development of Change Detection over Time

The history of change detection started with the emergence of remote sensing, and especially the first aerial photography taken in 1859 by Gaspard Félix Tournachon, also known as Nadar. Thereafter, the development of change detection is closely associated with military technology during World Wars I and II and the strategic advantage provided by the temporal information acquired by remote sensing. Civilian applications of change detection were developed following these events in the twentieth century, using mostly interpretation and analog means. However, the availability of such data to civilians varied around the world due to the military classification of imagery. While some countries in Europe provided this access following World War I, others in North America limited this access until the 1970s and 1980s.

The era of digital change detection started with the launch of Landsat-1 (initially known as the Earth Resources Technology Satellite) in July 1972. The regular acquisition of digital data on the Earth's surface in multispectral bands allowed scientists to get relatively consistent data over time and to characterize changes over a relatively large area for the first time. The continuity of that mission as well as the launches of numerous other missions drove the development of change detection techniques from that time.

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_7

However, the development of digital change detection techniques was limited by data processing technology capacities, and was therefore closely coupled to the development of computer technology. This technology evolved during the period from the 1960s, when only a few places in the world were equipped with (expensive) computers, to the present, when personal computers are fast and inexpensive enough to feasibly apply even complex algorithms and change detection techniques to satellite imagery. Computer technology also evolved from dedicated hardware to relatively user-friendly software specialized for image processing and change detection.

According to the published literature, algebraic techniques such as image differencing and image ratioing were the first techniques to be used to characterize changes in digital imagery during the 1970s [2]. These techniques are simple and fast to perform and are still widely used today. More complex techniques have been developed since then due to improvements in processing capacities but also the development of new theoretical approaches. Change detection analysis of the Earth's surface is currently a very active topic due to concerns about the consequences of global and local changes. This field of expertise is constantly progressing.

7.3 Overview of Methods

7.3.1 Changes on the Earth's Surface

The Earth's surface is changing constantly in many ways. First, the timescales at which changes can occur are very heterogeneous. They vary from those of catastrophic events (e. g., floods) to those of geological events (e. g., continental drift), which corresponds to a gradient between intermittent and continuous changes, respectively. Second, the spatial scales at which changes occur are also very heterogeneous, and vary from local events (e. g., road construction) to global changes (e. g., changes in the ocean water temperature). Due to this very large spatiotemporal range, the nature and extent of surface changes are complex to determine because they are interrelated and interdependent at different scales (spatial, temporal). Change detection is, therefore, a challenging task.

7.3.2 Imagery Characteristics Regarding Changes

Since the development of civilian remote sensing, there has been continuous and increasing coverage of the Earth by, for example, aerial photography or satellite imagery. This coverage is ensured by various sensors with various properties. First, various temporal resolutions (i. e., revisit times) and mission continuities allow coverage of every point on the Earth over timescales ranging from days to decades. Second, in terms of the spatial scale, imaging at various spatial resolutions (i. e., pixel size, scene size) provides coverage of every point on the Earth at submeter to kilometer resolution. Third, sensors are designed to observe the Earth's surface in various parts of the electromagnetic spectrum (i. e., spectral domains) at different spectral resolutions. In addition, the increasing use of unmanned aerial systems (UASs) for image acquisition is paving the way for new resolutions that have until now been difficult to access, such as millimetric spatial resolution or revisit times of less than a day. This diversity allows the characterization of a large spectrum of Earth surface elements and change processes. However, change detection is still limited by data availability and data consistency (i. e., multisource data).

7.3.3 Changes in Imagery

Changes in imagery between two dates translate into changes in radiance. Various factors can induce changes in radiance between two dates, such as changes in sensor calibration, solar angle, atmospheric conditions, seasons, or on the Earth's surface. The first premise of using imagery to detect changes on the Earth's surface is that these changes must result in a change in radiance. Second, the change in radiance due to a change on the Earth's surface must be large compared to the changes in radiance due to other factors. The challenge in performing change detection for the Earth's surface using imagery is to minimize these other factors. This is usually achieved by carefully selecting multidate imagery and applying preprocessing treatments.

7.3.4 Data Selection and Preprocessing

Data selection is a critical step in change detection studies. The acquisition period (i. e., season, month) of multidate imagery is an important parameter to consider in image selection because it is directly related to phenology, climatic conditions, and the solar angle. Careful selection of multidate images is therefore needed to minimize the effects of these factors. In vegetation change studies (i. e., performed over different years), for example, summer is usually used as the target period because of the relative stability of phenology, solar angle, and climatic conditions. The acquisition interval between multidate imagery is also important to consider. As mentioned before, surface changes of the Earth must cause enough of a change in radiance to be detectable. However, data selection is often limited by data availability and is usually a compromise between the targeted period, interval of acquisition, and availability. The cost of imagery is also a limiting factor in data selection.

However, careful data selection is usually not enough to minimize radiometric heterogeneity between multidate images. First, atmospheric conditions and solar angle differences usually need additional corrections; second, other factors such as sensor calibration or geometric distortions need to be considered. In change detection analysis, multidate images are usually compared on a pixel basis, so very accurate registrations need to be performed between images in order to compare pixels at the same locations. Misregistration between multidate images can cause significant errors in change interpretation [3]. The sensitivity of change detection approaches to misregistration is variable though. The minimization of radiometric heterogeneity (due to sources other than changes on the Earth's surface) can be achieved using different approaches, depending on the level of correction required and the availability of atmospheric data. Techniques such as dark object subtraction, relative radiometric normalization, and radiative transfer codes can be used.

7.3.5 Units of Analysis

As mentioned before, multidate images are usually compared on a pixel basis. However, the development of complex techniques in combination with the improvement in processing capacities has led to the use of several other units of analysis too. Table 7.1 provides an overview of five analytical units used in change detection studies.

This chapter presents change detection approaches using pixel-based examples. As shown in Table 7.1, there are other units of analysis that are also used, and approaches have been developed for each of those. However, since the pixel-based approaches presented in this chapter are generally applicable regardless of the unit of analysis [4], a detailed review of each category is not provided here. Systematic reviews are available for other units of analysis and for specific approaches such as object-based change detection (OBCD) [4, 5].

Table 7.1 Description of units of analysis used in change detection studies (adapted from [4])

Unit of analysis	Description
Pixel	Individual pixels comparison
Kernel	Group of pixels in a kernal filter or moving window
Image object	Image objects generated by image segmentation (independently or simultaneously in the time series)
Vector polygon	Polygons extracted from digital mapping or existing vector layers
Hybrid	Image objects generated from a pixel- or kernel-level comparison

7.3.6 Change Detection Methods

Summarized here are the most common methods used in change detection studies [1, 2, 4–9]. Most of these methods use image processing approaches that are applied to multidate satellite imagery. For the purposes of illustration, each method description is generally based on the pixel as the unit of analysis. It must be kept in mind that these methods are also applicable when using other units of analysis, as described in Sect. 7.3.5.

Image Differencing

This simple method is widely used and consists of subtracting a registered image acquired at one time from a registered image acquired at another time. For a pixel-by-pixel and band-by-band comparison, the method is defined as

$$Dx_{ij}^k = x_{ij}^k(t_2) - x_{ij}^k(t_1),$$

where Dx_{ij}^k is the difference between pixel value x located in row i and column j for band k between acquisition date 1 (t_1) and date 2 (t_2) [1].

If there is no change between the acquisition dates, the difference in pixel values will be 0, but, if change does occur, this difference will be positive or negative (Fig. 7.1).

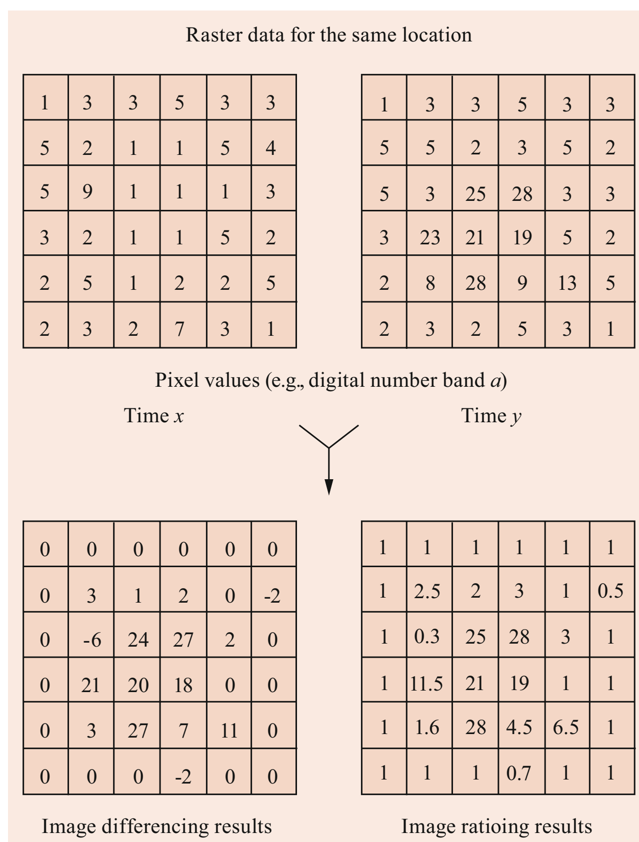


Fig. 7.1 Examples of the image differencing and image ratioing procedures

However, in practice, exact image registration and perfect radiometric corrections are never obtained for multidate images. Residual differences in radiance that are not caused by land cover changes are observed. Then, the challenge with this technique is to identify threshold values for change and no change in the resulting images. The standard deviation is often used as a reference value to select these thresholds. Different normalization, histogram-matching, and standardization approaches are used for multidate images to reduce scale- and scene-dependent effects on differencing results. The image differencing method is usually applied to single bands but can be also applied to processed data such as multidate vegetation indices or principal components.

Image Ratioing

This method is comparable to the image differencing method in terms of its simplicity and challenges. However, it is not as widely used. It is the ratio of registered images acquired at different times. For a pixel-by-pixel or band-by-band comparison, the method is defined as

$$Rx_{ij}^{k} = \frac{x_{ij}^{k}(t_1)}{x_{ij}^{k}(t_2)},$$

where Rx_{ij}^{k} is the ratio between pixel value x located in row i and column j for band k between acquisition date 1 (t_1) and date 2 (t_2) [1].

Changes are represented by pixel value ratios that are higher or lower than 1 (Fig. 7.1). No change is indicated by a ratio of 1. In practice, for the same reasons as in image differencing, the challenge of this technique is in selecting the appropriate threshold value between change and no change. This technique is often criticized because the nonnormal distribution of the results limits the validity of threshold selection using the standard deviation of the resulting pixels.

Postclassification

This method is also commonly referred to as *delta classification*. It is widely used and easy to understand. Two images acquired at different times are independently classified and then compared. Ideally, similar thematic classes are produced for each classification. Changes between the two dates can be visualized using a change matrix indicating, for both dates, the number of pixels in each class (Fig. 7.2). This matrix allows one to interpret the changes that occurred for a specific class. The main advantage of this method is the minimal impact of radiometric and geometric differences between multidate images. However, the accuracy of the final result is the product of the accuracies of the two independent classifications (e. g., a final accuracy of 64% is obtained for two independent classification accuracies of 80%).

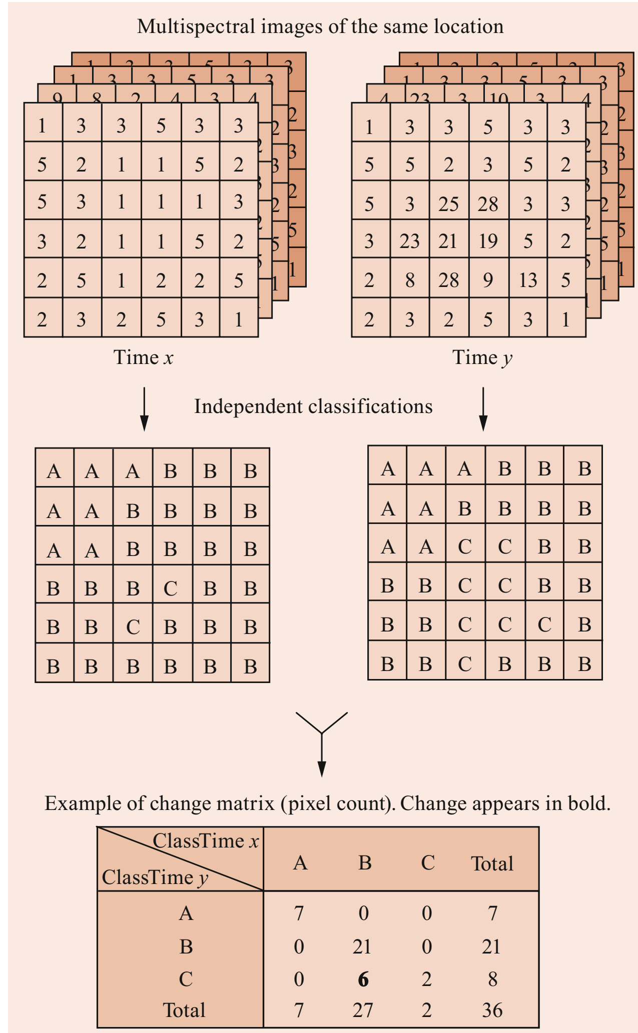


Fig. 7.2 Example of a postclassification procedure

Direct Multidate Classification

This method is also referred to as *composite analysis*, *spectral-temporal combined analysis*, *spectral-temporal change classification*, *multidate clustering*, or *spectral change pattern analysis*. Multidate images are combined into a single dataset for which classification is performed (Fig. 7.3). The areas with changes are expected to present different statistics (i. e., distinct classes) compared to the areas with no changes. This approach can be implemented unsupervised or supervised, and necessitates only one classification procedure. However, this method usually produces numerous classes corresponding to spectral changes within each image but also temporal changes between images. The interpretation of results is often complex and requires a good knowledge of the study area. Combined approaches using principal component analysis or machine learning algorithms [4] can be performed to reduce data dimensionality or the coupling between spectral and temporal change, respectively.

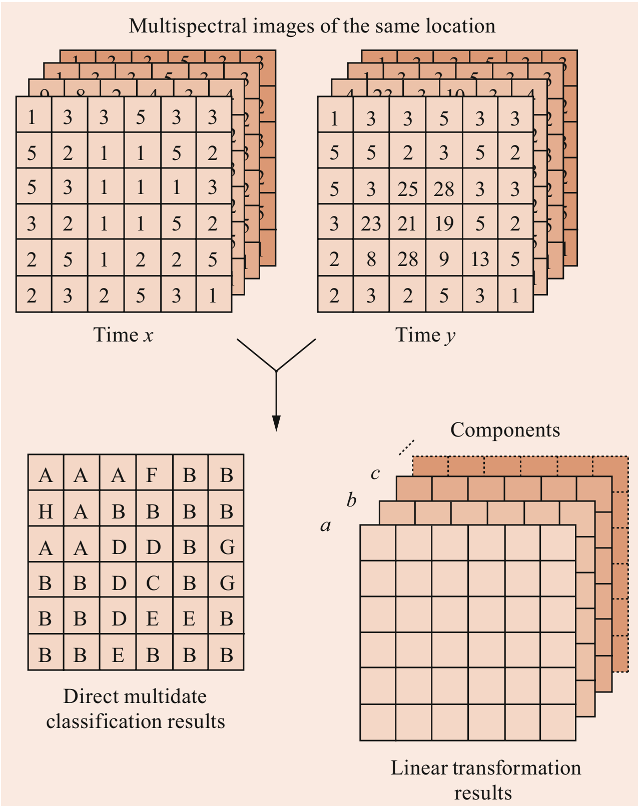


Fig. 7.3 Example of direct multidate classification and linear transformation procedures

Linear Transformation

This approach includes different techniques with the same theoretical basis. Principal component analysis (PCA) and tasseled-cap transformation are the most common ones [7]. Linear transformations are often used to reduce spectral data dimensionality by creating fewer new components. The first few components contain most of the variance in the data and are uncorrelated. When used for change detection purposes, linear transformations are performed on multidate images that are combined into a single dataset (Fig. 7.3).

After performing a PCA, unchanged areas are mapped in the first component (i. e., information common to multidate images), whereas areas with changes are mapped in the other components (i. e., information unique to either of the dates). Usually, the PCA is calculated from a variance/covariance matrix. However, a standardized matrix (i. e., a correlation matrix) is also used. PCA is scene dependent and results can be hard to interpret. For example, a change analysis performed on one part of an image will provide different results for the same area compared to an analysis performed on the whole image. The challenging steps are to label changes from principal components and to select the thresholds between changed and unchanged areas. A good knowledge of the study area is required.

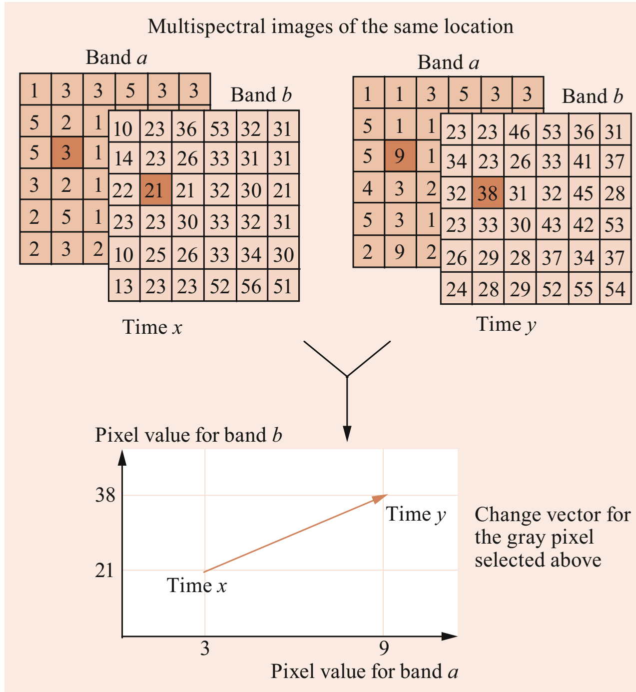


Fig. 7.4 Example and principle of the change vector procedure

The tasseled cap is also a linear transformation. However, unlike PCA, it is independent of the scene. The new component directions are selected according to predefined spectral properties of vegetation. Four new components are computed and oriented to enhance brightness, greenness, wetness, and yellowness. Results are also difficult to interpret, and change labeling is challenging. For example, defining changed areas involves the selection of thresholds based on the analyst's preferences, which is relatively subjective; the threshold may vary according to the expertise of the analyst. Unlike PCA, the tasseled-cap transformation for change detection requires accurate atmospheric calibration of multidate imagery.

Other transformations such as multivariate alteration detection or Gram–Schmidt transformation have also been developed but are used to a lesser extent.

Change Vector Analysis

This approach is based on the spatial representation of change in a spectral space. When a pixel undergoes a change between two dates, its position in n-dimensional spectral space is expected to change. The change is represented by a vector (Fig. 7.4) that is defined by two factors: the direction, which provides information about the nature of the change, and the magnitude, which provides information about the level of change. This approach has the advantage that it can process any number of spectral bands concurrently. It also provides detailed information about the change. The challenging steps are to define the threshold of magnitude that discriminates

7

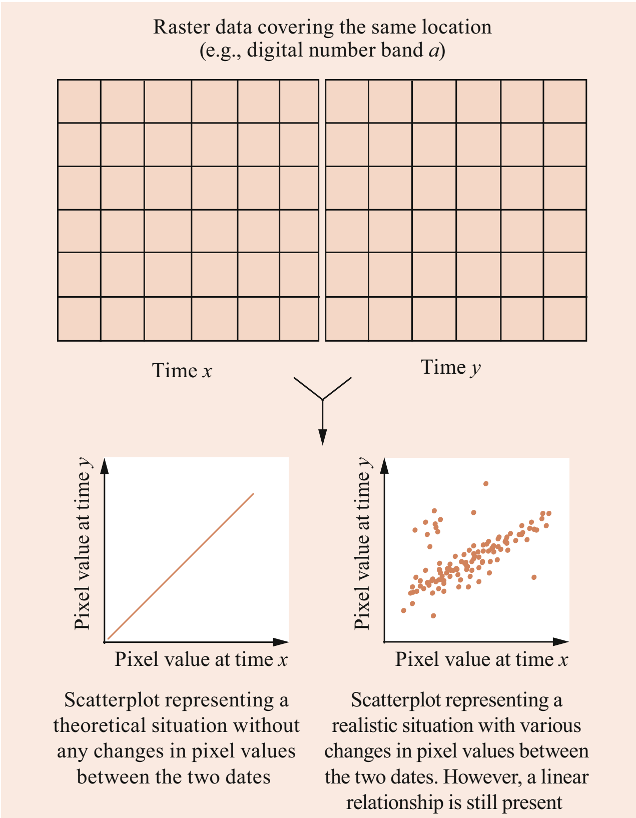


Fig. 7.5 Example and principle of the image regression procedure

between change and no change and to interpret the vector direction in relation to the nature of the change. For example, the analysis of high-dimensional (e. g., multi- or hyperspectral) datasets produces high-dimensional vectors that are difficult to interpret. In addition, different types of changes can be characterized by similar values of direction and magnitude, which can lead to confusion during data interpretation. This approach is often performed on transformed data using methods such as the tasseled-cap transformation.

Image Regression

This approach assumes that there is a linear relationship between the pixel values of the same area at two different times. This implies that the majority of the pixels did not change between the two dates (Fig. 7.5). A regression function that best describes the relationship between the pixel values for each spectral band on the two dates is developed. The residuals of the regression are considered to represent the areas with changes. Changes can also be detected by applying the image differencing method to the predicted image values at date 2 (computed using the regression function) and image values at date 1, i. e.,

$$Dx_{ij}^{k} = \hat{x}_{ij}^{k}(t_2) - x_{ij}^{k}(t_1),$$

where $\hat{x}_{ij}^{k}(t_2)$ is the predicted value of pixel x located in row i and column j for band k at date 2 (t_2) [1].

This method has the advantage of reducing the impact of radiometric heterogeneity (i. e., due to the atmosphere, sun angle, sensor calibration) between multidate images. However, the challenging steps are to select an appropriate regression function and to define the threshold between changed and unchanged areas.

Multitemporal Spectral Mixture Analysis

Spectral mixture analysis is based on the premise that a pixel reflectance value can be computed from the individual values of its constituent elements (i. e., endmembers) weighted by their respective proportions via

$$DN_c = \sum_{i=1}^{n} F_i \, DN_{i,c} + E_c$$

with the constraints

$$\sum_{i=1}^{n} F_i = 1 \quad \text{and} \quad 0 \le F_i \le 1,$$

where DN_c is the digital number of the pixel value in channel c, F_i is the fraction of endmember i; $DN_{i,c}$ is the digital number value of endmember i in channel c, n is the number of endmembers, and E_c is the error in the estimate for channel c. This approach aims to solve the equation and find the fraction of each endmember in each pixel. The root mean square error is calculated for each pixel and minimized to obtain the best model. Endmember fractions must be between 0 and 1 and sum to 1 [10].

This case assumes linear mixing of these components. The method allows the retrieval of subpixel information (i. e., surface proportions of endmembers), and can be used for change detection purposes by performing separate analyses and comparing results at different dates (Fig. 7.6). The advantage of this method is that it provides precise and repeatable results, although it is challenging to select suitable endmembers.

Combined Approaches

The previous techniques represent the most common approaches used for change detection. They can be used individually, but are often used in combination or with other image processing techniques to provide more accurate results. Numerous combinations can be used, but they will not be described here. Some of them include vegetation indices and image differencing, change vector analysis and principal component analysis, direct multidate classification and principal component analysis, multitemporal spectral analysis and image differencing, and image enhancement and postclassification.

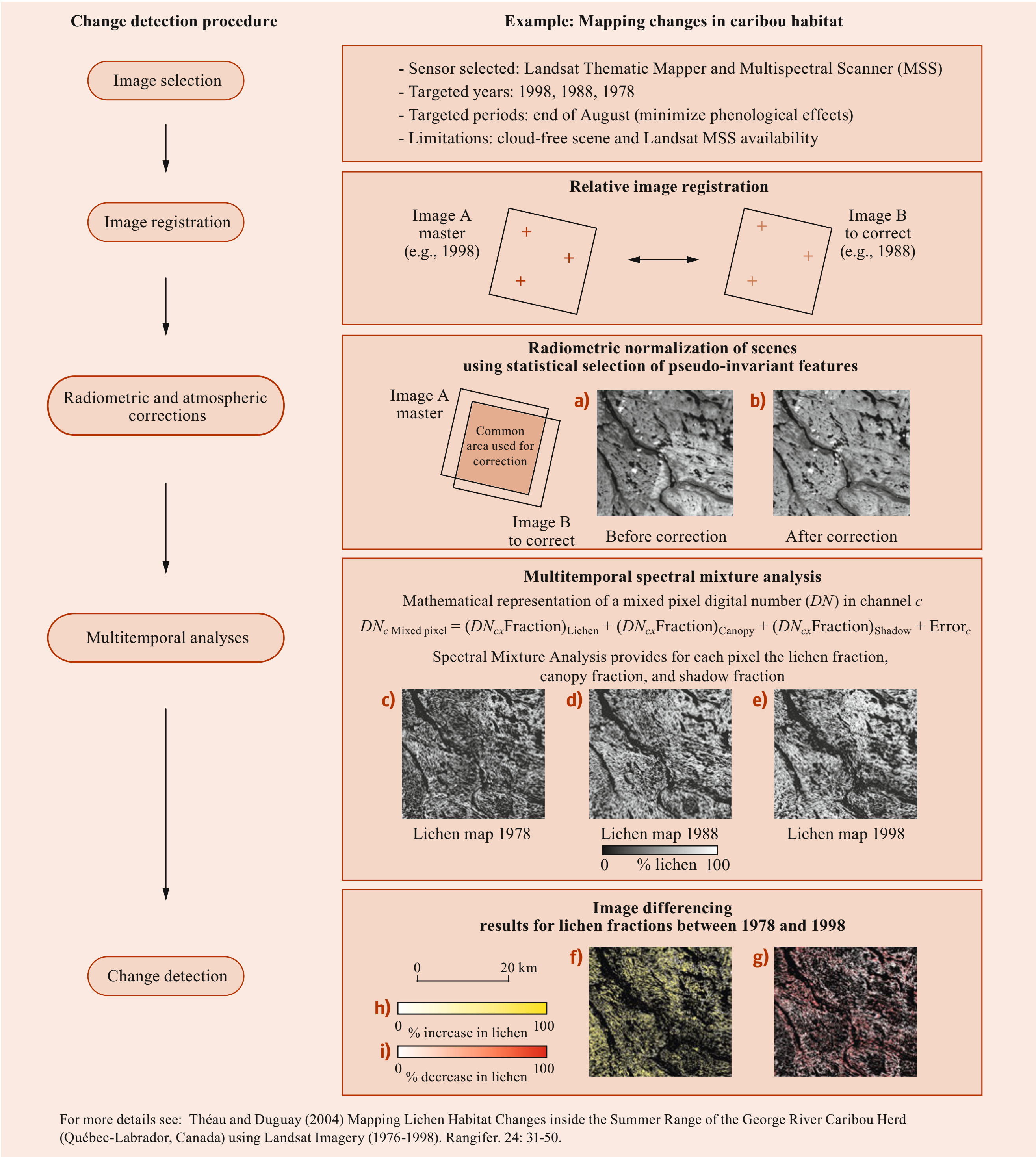


Fig. 7.6 Example of a change detection procedure. Case study of the mapping of changes in caribou habitat using multitemporal spectral mixture analysis

Example of Change Detection Analysis (Mapping Changes)

The George River Caribou Herd (GRCH), located in northeastern Canada, increased from about 5000 in the 1950s to about 700 000 head in the 1990s. This has led to an overutilization of summer habitat, resulting in degradation of the vegetation cover. This degradation has had a direct impact on health problems observed in the caribou (*Rangifer tarandus*) population over the last few years, and may also have contributed to the recent decline of the GRCH (404 000 head in 2000–2001). Lichen habitats are good indicators of caribou herd activity because of their sensitivity to overgrazing and overtrampling, their widespread distribution over northern territories, and their influence on herd nutrition. The herd range covers a very large territory which is not easily accessible. As a result, field studies over the whole territory are limited and aerial surveys cannot be conducted frequently. Satellite imagery offers the synoptic view and temporal resolution necessary for mapping and monitoring caribou habitat. In this example, a change detection approach using Landsat imagery was used. The procedure was based on spectral mixture analysis and produced maps showing the lichen proportion inside each pixel. The procedure was applied to multidate imagery to monitor the spatiotemporal evolution of the lichen resource over the past three decades. It gave new information about the habitat used by the herd in the past, which was very useful to better understand caribou population dynamics. Figure 7.6 summarizes the approach used in this study and illustrates the steps typical of a change detection procedure.

7.4 Typical Applications

The Earth's surface is changing constantly in many ways. Changes occur at various spatial and temporal scales in numerous environments. Change detection techniques are employed for different purposes such as research, management, or business. Change monitoring using GIS and remote sensing is therefore used in a wide field of applications [2, 11–23]. A nonexhaustive list of key applications is presented below.

7.4.1 Forestry and Conservation

- Deforestation (e. g., clearcut mapping, regeneration assessment, carbon storage)
- Fire monitoring (e. g., delineation, severity, detection, regeneration)
- Logging planning (e. g., infrastructure, inventory, biomass)
- Herbivory (e. g., insect defoliation, grazing)
- Habitat fragmentation (e. g., landcover changes, heterogeneity, ecological integrity)

7.4.2 Agriculture and Rangelands

- Crop monitoring (e. g., growing, biomass)
- Invasive species (e. g., detection, distribution)
- Soil moisture condition (e. g., drought, flood, landslides)
- Desertification assessment (e. g., bare ground exposure, wind erosion).

7.4.3 Urban

- Urban sprawl (e. g., urban mapping)
- Transportation and infrastructure planning (e. g., landcover use)
- Georisk (e. g., earthquakes, volcanoes, subtle deformation, structural integrity).

7.4.4 Ice and Snow

- Navigation route (e. g., sea ice motion)
- Infrastructure protection (e. g., flooding monitoring, avalanche risk)
- Glacier and ice sheet monitoring (e. g., motion, melting)
- Permafrost monitoring (e. g., surface temperature, tree line).

7.4.5 Ocean and Coastal

- Water quality (e. g., temperature, productivity)
- Aquaculture (e. g., productivity)
- Intertidal zone monitoring (e. g., erosion, vegetation mapping)
- Oil spill (e. g., detection, oil movement).

7.5 Probable Future Directions

In the past few decades, we have observed a constant increase in remotely sensed data availability. The launches of numerous satellite sensors as well as a reduction in product costs can explain this trend. It will continue in the future at an even faster pace due to the exponential use of imagery acquired by UASs. Access to constantly growing archive content also provides the potential to develop more change detection studies in the future. Long-term missions such as Landsat, SPOT (Satellite pour l'Observation de la Terre), AVHRR (the Advanced Very-High-Resolution Radiometer) have provided continuous data for over 30–40 years. Although radiometric heterogeneity between sensors represents a serious limitation of time series analysis, these data are still very useful for long-term change studies. They are particularly suitable for temporal trajectory analysis, which usually involves the temporal study of indicators (e. g., vegetation indices, surface temperature) at the global scale.

Moreover, as mentioned before in Sect. 7.2, the development of change detection techniques is closely linked with the development of computer technology and data processing capacities. In the future, these fields will still evolve in parallel, so new developments in change detection are expected to occur with the further development of computer technology.

The development and application of new image processing methods and geospatial analysis are also expected in the coming decades. Artificial intelligence systems as well as knowledge-based expert systems and machine learning algorithms represent new alternative techniques for change detection studies [6]. These techniques have gained considerable attention in the past few years and are expected to be increasingly applied in change detection approaches in the future, especially in the context of multiple-resolution imagery processing (particularly when images with very high spatial resolution are present). One of the main advantages of these techniques is that they allow the integration of existing knowledge with nonspectral information on the scene content (e. g., socioeconomic data, shape, and size data). With the increasing interest in using integrated approaches such as coupled human–environment systems, these developments look promising.

The recent integration of change detection and spatial analysis modules into most GIS software also represents a big step towards integrated tools for the study of changes on the Earth's surface. This integration also includes improving the compatibility between image processing software and GIS software. More developments that will provide new tools for integrating multisource data more easily (e. g., digital imagery, hard maps, historical information, and vector data) are expected in the future.

References

1. Singh, A.: Digital change detection techniques using remotely-sensed data. Int. J. Remote Sens. **10**, 989–1003 (1989)
2. Lunetta, R.S., Elvidge, C.D.: Remote Sensing Change Detection: Environmental Monitoring Methods and Applications. Ann Arbor Press, Chelsea (1998)
3. Weber, K., Théau, J., Serr, K.: Effect of coregistration error on patchy target detection using high-resolution imagery. Remote Sens. Environ. **112**, 845–850 (2008)
4. Tewkesbury, A.P., Comber, A.J., Tate, N.J., Lamb, A., Fisher, P.F.: A critical synthesis of remotely sensed optical image change detection techniques. Remote Sens. Environ. **160**, 1–14 (2015)
5. Chen, G., Hay, G.J., Carvalho, L.M.T., Wulder, M.A.: Object-based change detection. Int. J. Remote Sens. **33**, 4434–4457 (2012)
6. Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., Lambin, E.: Digital change detection methods in ecosystem monitoring: a review. Int. J. Remote Sens. **25**, 1565–1596 (2004)
7. Lu, D., Li, G., Moran, E.: Current situation and needs of change detection techniques. Int. J. Image Data Fusion **5**, 13–38 (2014)
8. Mas, J.-F.: Monitoring land-cover changes: a comparison of change detection techniques. Int. J. Remote Sens. **20**, 139–152 (1999)
9. Bruzzone, L., Bovolo, F.: A novel framework for the design of change-detection systems for very-high-resolution remote sensing images. Proc. IEEE **101**, 609–630 (2012)
10. Adams, J.B., Smith, M.O., Gillespie, A.R.: Simple models for complex natural surfaces: a strategy for the hyperspectral era of remote sensing. In: Proc. Int. Geosci. Remote Sens. Symp. (IGARSS'89)/12th Can. Symp. Remote Sens., vol. 1, pp. 16–21. (1989)
11. Ellis, E.C., Wang, H., Xiao, H.S., Peng, K., LIu, X.P., Li, S.C., Ouyang, H., Cheng, X., Yang, L.Z.: Measuring long-term ecological changes in densely populated landscapes using current and historical high resolution imagery. Remote Sens. Environ. **100**, 457–473 (2006)
12. Willis, K.S.: Remote sensing change detection for ecological monitoring in United States protected areas. Biol. Conserv. **182**, 233–242 (2015)
13. Kennedy, R.E., Townsend, P.A., Gross, J.E., Cohen, W.B., Bolstad, P., Wang, Y.Q., Adams, P.: Remote sensing change detection tools for natural resource managers: understanding concepts and tradeoffs in the design of landscape monitoring projects. Remote Sens. Environ. **113**, 1382–1396 (2009)
14. Dong, L., Shan, J.: A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. ISPRS J. Photogramm. **84**, 85–99 (2013)
15. Patino, J.E., Duque, J.C.: A review of regional science applications of satellite remote sensing in urban settings. Comput. Environ. Urban. Syst. **37**, 1–17 (2013)
16. Hansen, M.C., Loveland, T.R.: A review of large area monitoring of land cover change using Landsat data. Remote Sens. Environ. **122**, 66–74 (2012)
17. Malinverni, E.S.: Change detection applying landscape metrics on high remote sensing images. Photogramm. Eng. Remote. Sens. **77**, 1045–1056 (2011)
18. Goetz, S., Dubayah, R.: Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. Carbon Manag. **2**, 231–244 (2011)
19. Canada Centre for Mapping and Earth Observation (CCMEO):. http://www.nrcan.gc.ca/earth-sciences/geomatics/10776, Accessed 12.2017
20. ESA: Observing the earth. http://www.esa.int/esaEO/, Accessed 12.2017
21. Global Change Master Directory:. https://gcmd.nasa.gov/learn/index.html, Accessed 12.2017
22. IHDP: International human dimensions programme on global environmental change. http://www.ihdp.org/, Accessed 12.2017
23. WCRP: World Climate Research Programme. https://www.wcrp-climate.org/, Accessed 12.2017

Jérôme Théau Jérôme Théau is Professor at Université de Sherbrooke (Canada). He has a background in wildlife and habitat management, and remote sensing. He has been working for more than 20 years on spatiotemporal characterization of ecosystems. Some of his main interests include the study of wildlife habitats and of landscape fragmentation from a conservation point of view using satellite imagery, GIS, and drones.

7

Geodesy

8

Matthias Becker

Contents

Abstract

Geodesy is the basis for all geographic information system (GIS) applications, as it provides all the information required to describe the location of a point at or close to the Earth. In this chapter, the basic definitions, quantities, and mathematical relations used in geodesy are described. The chapter should provide a good understanding of reference frames, coordinates, and height systems, their variations over time, and their relations to plane coordinates. It also includes a review of the importance of the Earth's gravity field and basic methods to determine coordinates.

Keywords

Reference Systems · Coordinate Systems · Height Systems · Coordinate Transformations · Time Scales · Geopotential · Geoid · GNSS

8.1 Basics

Geodesy is closely related to the natural sciences as well as to the technical sciences. Geodetic calculations strictly utilize the *Système International d'Unités* (SI) system of physical units [m–kg–s] and can, to a large extent, be represented by Euclidian geometry and Newtonian mechanics. Geodesy deals with the determination of the size and shape of the Earth, its gravity field, and the geometric coordinates of surface or other points related to the Earth. The availability of artificial Earth satellites and advances in space geodesy have tremendously facilitated this task. For the first time, it became possible to derive the precise shape of the Earth, its gravity field, and well-defined global reference frames to monitor coordinates and changes in them over time. Tasks that geodesists worked on for centuries are now completed within weeks. Presently, the Earth is being covered with a network of thousands of continuous global satellite navigation system (GNSS) stations [1–3] that continuously record variations in coordinates at the

M. Becker (✉)
Institute of Geodesy, Faculty of Civil and Environmental Engineering, Technical University of Darmstadt
Darmstadt, Germany
e-mail: becker@psg.tu-darmstadt.de

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_8

millimeter level. GPS, having been the first of four global navigation satellite systems (GNSSs) to become operational, is still the main tool used for positioning, in particular for geographic information system (GIS) applications. However, Multi-GNSS receivers that also use the Russian GLONASS, Chinese BeiDou (BDS), and European Galileo GNSS 9 are becoming more and more popular, as the use of several systems at once increases accuracy, reliability, and integrity [4]. Coordinates can be measured in a global, common, and homogeneous reference system to accuracies ranging from tens of meters in navigation down to the submillimeter level in deformation analysis.

Geometric monitoring of regional surface deformations is enabled by remote sensing satellites equipped with synthetic aperture radar (SAR) and interferometric differential SAR. These techniques allow the derivation of digital terrain models at the sub-meter level and the detection of variations in terrain at the millimeter level. Other geometric techniques that are applied include airborne radar and laser-based lidar.

The geometric information obtained from GNSS, radar, lidar, and photogrammetry has to be supplemented by gravity field information [5] if the vertical component or a physical, gravity-field-related height component is of interest. Today, gravity field information comes from a combination of satellite mission data and high-resolution terrestrial gravity observations. It has a global accuracy of a few centimeters to decimeters if used for the determination of the physical reference surface for height, i.e. the geoid (commonly referred to as *mean sea level*). Combining the geoidal height and the GNSS height leads to the vertical coordinate that is relevant in GIS and many other surveying and mapping applications, commonly referred to as *the height*. At the present level of accuracy, time-dependent geoid changes can be monitored at the millimeter level averaged over regions of about 500 km^2. These changes may be caused by gravitational changes due to mass redistribution, such as melting icecaps or glaciers or other global and local effects.

The third pillar of geodesy besides geometry and gravity—the determination of the Earth's rotation and motion in space—is not considered in the framework of geographic information systems here; a summary of this topic can be found in [6].

The basic task of geodesy is the definition and realization of coordinate systems and their interrelations that allow the continuously changing Earth to be described. Through the collection and administration of all data describing the geometric and physical structure of the Earth's surface, users can apply this information to produce GIS systems that archive, display, and utilize these data for all types of applications. The proper use of spatial data requires an understanding of these basic principles and relationships, as this allows us to correctly assess quality of and uncertainty in data and to correctly handle geodetic data.

8.2 Concepts

Geodesy makes use of a number of surfaces that have to be clearly distinguished (Fig. 8.1). These are the solid Earth surface (i. e., the topography), the reference ellipsoid, and equipotential surfaces such as the geoid or a local level surface. The position of a point in space can be described purely geometrically by three-dimensional Cartesian coordinates referenced, for example, to the center of mass of the Earth.

However, describing the spatial relation of the position of this point to features on the Earth's surface in terms of latitude, longitude, and ellipsoidal height is more appropriate and informative. As it is purely geometric, the introduction of a sphere or (better) a reference or mean Earth ellipsoid allows the separation of the horizontal position from the vertical position above the ellipsoid. A mean sphere of radius $R = 6371.0$ km may be sufficient for some applications, but the use of an ellipsoid is more appropriate and still simple enough to handle in computations. The flattening (i. e., the oblateness) of an ellipsoid fitted optimally to the Earth results in a difference of 23 km between the equatorial and the polar axes.

When aiming for an even better approximation of the Earth's shape, the concept of level surfaces plays a major role. Level surfaces are defined as those that are normal everywhere to the direction of the plumb line. On a level surface, therefore, water cannot flow, and any liquid will be at rest if it is part of a level surface. In particular, the undisturbed surface of the oceans—the mean sea surface—is very close to being a level surface. The level surface that is the best fit to the mean sea surface is called the *geoid*, which extends inside the Earth's crust and serves as a reference for height determinations. The geoid is a complicated surface that cannot be described analytically and therefore cannot be used for computations. It is described by the distance to the best-fitting ellipsoid, which deviates by up to about 100 m from the geoid (Fig. 8.2). This vertical separation of ellipsoid and geoid that determines the shape of the geoid is computed globally from models of the geopotential, i. e., Earth gravity models (EGM), or locally from more precise regional and local geoidal models.

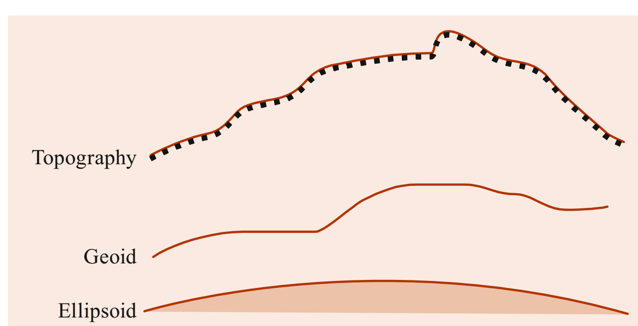


Fig. 8.1 Main geodetic surfaces

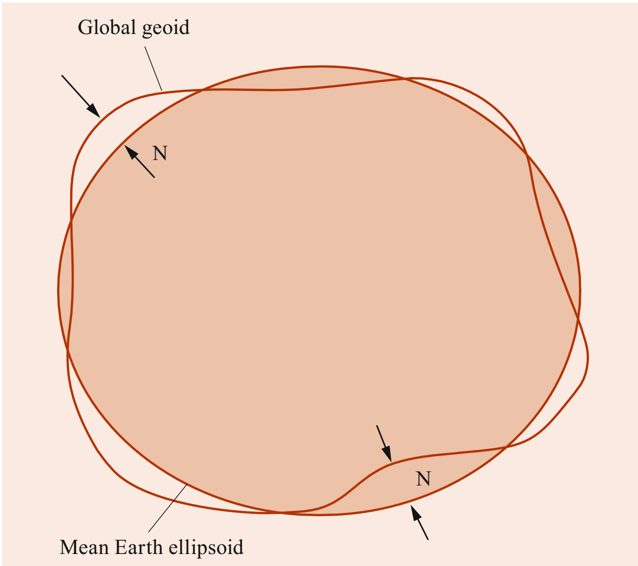


Fig. 8.2 Mean Earth ellipsoid and global geoid

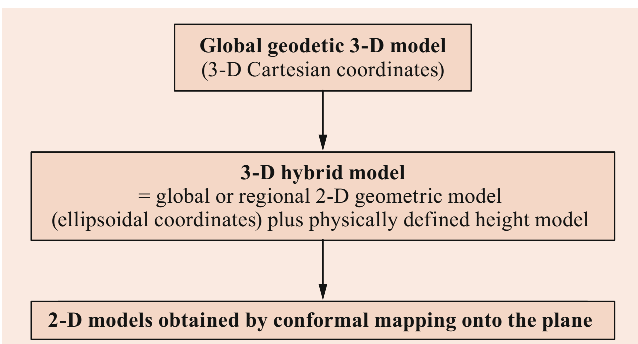


Fig. 8.3 The hierarchy of the three layers of coordinate systems

When using coordinates in a GIS, we can distinguish conceptually between the three layers illustrated in Fig. 8.3. The first of these corresponds to the global geodetic three-dimensional (3-D) Cartesian coordinates for a unique description in geometry space. The next level is the use of a global or regional two-dimensional (2-D) surface model, such as the ellipsoid, that allows the separation of the vertical coordinate and the introduction of a level surface such as the geoid to work with physically defined height models. The third layer contains 2-D models that are obtained from the conformal projection of the curvilinear coordinates onto the plane, e. g., the most commonly used Universal Transverse Mercator (UTM) coordinates.

8.3 Reference Systems and Reference Frames

This section describes the basic reference systems and their realization in the form of reference frames. Reference systems are maintained by the International Earth Rotation and Reference System Service [7] based on the actual standards

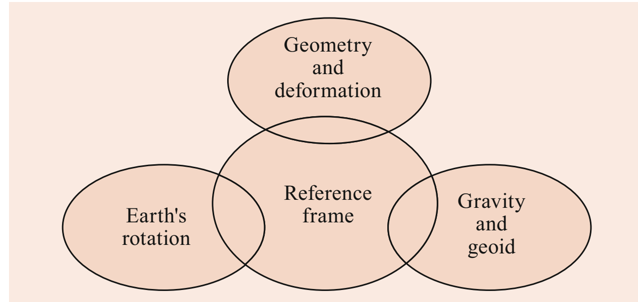


Fig. 8.4 The three pillars of geodesy linked together by the reference system

as defined by the International Association of Geodesy (IAG) and the International Astronomical Union (IAU). Fortunately, due to standardization and the predominant use of GNSS or other satellite techniques, virtually all actual positioning is based on these systems, and coordinates are given in the International Terrestrial Reference Frame (ITRF) [8]. Current regional or national systems are mostly based on the ITRF or are in the process of being updated to it, which highlights the utmost importance of the ITRF and the need for a proper definition and understanding. The celestial system [9] is as important, although not directly visible to the user of positioning services. In addition, we will introduce the World Geodetic System 1984 (WGS 84) due to its widespread use and the Geodetic Reference System 1980 (GRS80) due to its importance in mapping and national surveying.

As the Earth is rotating and satellites revolve in an inertial space, a superior system, the celestial reference system, must be used. This constitutes an inertial frame of reference in which any body initially at rest will remain at rest indefinitely, or in which a moving body moves in a straight line with constant speed indefinitely; in other words, it is free from any inertial forces. It can be defined as a frame of reference in which Newton's laws of motion apply exactly. Terrestrial, Earth-fixed systems are not inertial because they revolve around the Sun and rotate with the Earth, so virtual forces such as the Coriolis force and the centrifugal force have to be taken into account. Reference systems are constructed from observations by geodetic space techniques, and linking them allows the unique realization of both the celestial inertial system and the terrestrial system. In Fig. 8.4, the three pillars of geodesy are shown together with the observation techniques that are involved. Geometric observations, observations of the Earth's rotation, and gravity-field observations are combined to maintain the reference frame.

8.4 Coordinate Reference System

Coordinate reference systems (CRSs) are a combination of at least one coordinate system together with its spatial datum. In a CRS, positions or locations of geographic information

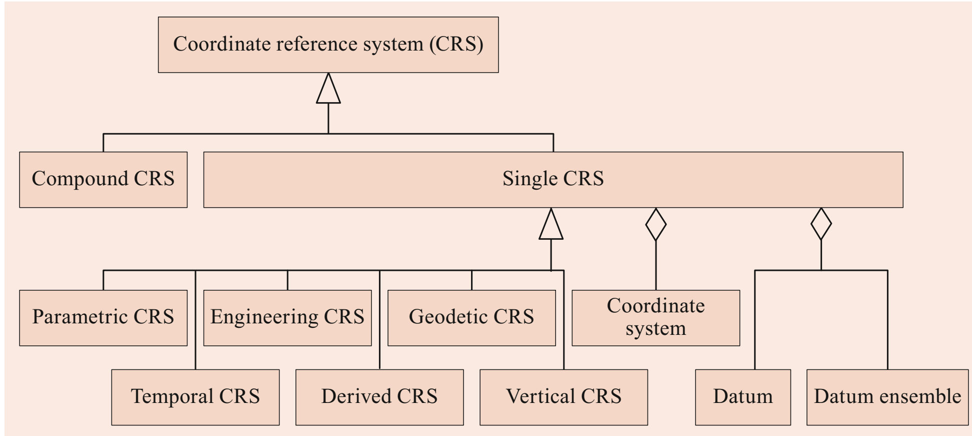


Fig. 8.5 Schema for the definition of a CRS (source: ISO 19111:2019 on spatial referencing by coordinates)

are described by coordinates. In a GIS [10], the schema for the definition of a CRS contains two different elements: the datum and the coordinate system (Fig. 8.5). The datum defines how the CRS is related to the Earth: the position of the origin, the scale, and the orientation of coordinate axes, e. g., ED50 (European Datum 1950) and ETRS89 (European Terrestrial Reference System 1989). A geodetic datum, in addition, includes the parameters of a reference ellipsoid. A vertical datum defines the reference potential of physical heights (Sect. 8.5). The datum may also be a local engineering datum. The coordinate system describes how the coordinates are expressed in the specified datum, e. g., as Cartesian coordinates, ellipsoidal coordinates, or coordinates of a map projection such as the UTM. The coordinate system—the mathematical part of the coordinate reference system (CRS)—is a set of rules, e. g., projection equations, for specifying how coordinates are to be assigned to points. The list of coordinates in a specified CRS constitutes the coordinate reference frame (CRF).

8.4.1 Coordinate Systems and Coordinate Types

In order to specify a location, three coordinates are required, along with a time stamp (as material points may be subject to motion). The methods and concepts that are used to fix a point in space are called the coordinate system. They are defined by conventions. The most common conventional coordinate system is the orthogonal system of Cartesian coordinates (Fig. 8.6)

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + x_3\boldsymbol{e}_3 = \sum_{i=1}^{3} x_i\boldsymbol{e}_i ,$$

$$\text{where} \quad e_i \cdot e_j = \delta_{\text{ij}} = \begin{cases} 0 & \text{if} \quad i \neq j \\ 1 & \text{if} \quad i = j \end{cases} . \tag{8.1}$$

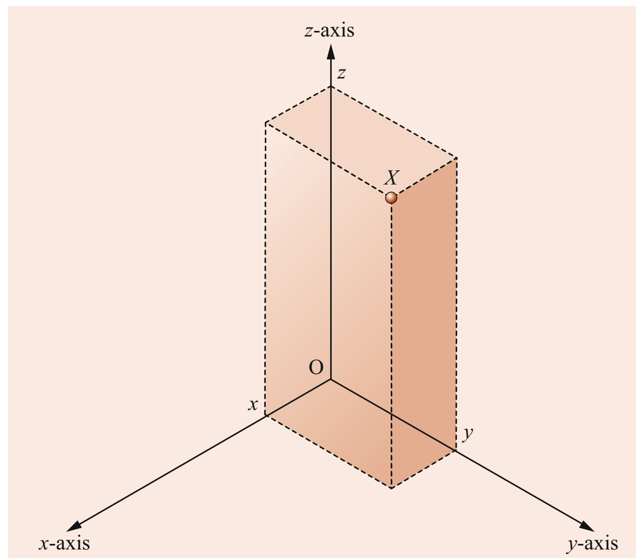


Fig. 8.6 Three-dimensional Cartesian coordinates

In three-dimensional geodetic applications, the components for the three axes are labeled x, y, and z, thus

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} . \tag{8.2}$$

The other main system of orthogonal coordinates used in geodesy is that of the geodetic surface coordinates of an ellipsoid of rotation. These are the ellipsoidal coordinates geodetic latitude φ, geodetic longitude λ, and ellipsoidal height h (Fig. 8.7). φ is the angle between the equatorial $(x,\ y)$-plane and the normal to the ellipsoid, λ is the angle between the $(x,\ z)$-plane located at the Greenwich mean meridian and the normal to the ellipsoid, and h is the normal distance to the ellipsoid. The size and figure of the ellipsoid are defined by the major and the minor axes a and b. The flattening f, the first and second eccentricities e and e',

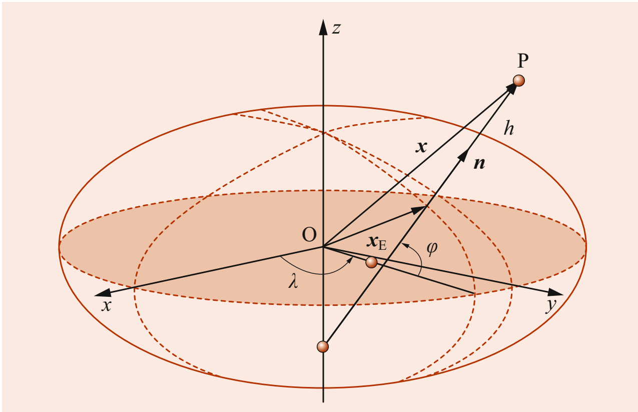


Fig. 8.7 Ellipsoidal coordinates and position vectors

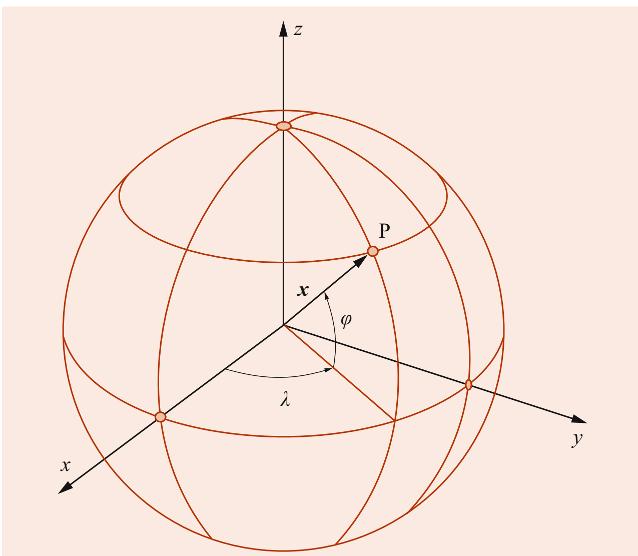


Fig. 8.8 Earth-fixed terrestrial system with geographic coordinates

and the polar radius of curvature c are derived constants that are used in calculations with ellipsoidal coordinates, e. g., in conversions from and to Cartesian coordinates (Sect. 8.8.1, Fig. 8.13). In general, the term *geodetic coordinates* is associated specifically with the ellipsoid used for large- and medium-scale mapping and in geodesy. The term *geographic coordinates* is more general and refers to spherical coordinates (Fig. 8.8) that are used for small-scale mapping or approximations to the ellipsoid.

An important distinction has to be made for natural or astronomical coordinates. These describe the direction of the plumb line. At each point they give the normal direction to the associated equipotential surface, i. e., the local zenith direction. Astronomical coordinates are dependent on the irregularities in the local gravity field, but may differ from the ellipsoidal latitude and longitude by up to 30′ or more in mountainous regions.

8.4.2 International Celestial Reference System and Frame

The celestial reference system is a conventional system that constitutes an inertial space-fixed system [11]. Its origin is located at the barycenter (center of mass) of the solar system. To define the coordinate axis, some further explanation of the Earth and its rotational motion is needed. The rotational axis of the Earth, or more precisely its angular momentum axis, is used as the x_3 axis of a Cartesian system. The position of this reference axis is called the Celestial Ephemeris Pole (CEP). As the rotation vector of the Earth oscillates for a number of reasons, a specific date (J2000.0, Sect. 8.3) at which the direction of the rotational axis is used has to be defined. The x_1 axis of the celestial system points towards the vernal equinox; it is specified by the direction of the intersection of the equatorial plane of the Earth and the ecliptic. The x_2 axis completes the orthogonal system of the International Celestial Reference System (ICRS).

The ICRS is realized by very-long-baseline interferometry (VLBI) estimates of the equatorial coordinates of a set of extragalactic compact radio sources: the International Celestial Reference Frame (ICRF). By appropriately modeling VLBI observations in the framework of general relativity, the directions of the CEP and the vernal equinox are maintained fixed relative to this selected set of precise coordinates of quasars. The catalog of stars used for optical astronomical observations (the *Fundamentalkatalog* FK5 or, presently, the HIPPARCOS catalog [12] is aligned to the ICRF and provides the primary realization of the ICRS at optical wavelengths).

The relation between the celestial and the terrestrial frames is of importance if astronomical positioning methods that use star cameras or GNSS orbit determination are required in the framework of navigation or positioning.

8.4.3 International Terrestrial Reference System and Frame

For all tasks that relate to the Earth's surface or the region close to it, an Earth-fixed system that rotates with the Earth is appropriate. This, again, is a conventional system that is geocentric, with the center of mass defined for the whole Earth, including the oceans and atmosphere. The Cartesian axis x_3 is attached to the mean rotational axis of the Earth [13]. The position of this mean axis on the Earth's surface is called the Conventional International Origin and was initially given by the Bureau International de l'Heure (BIH) orientation at 1984.0. The direction of the x_1 axis is the BIH 1984.0 mean meridian of Greenwich, and x_2 completes the orthogonal system. The latter two axes define the terrestrial plane of the equator orthogonal to the mean rotational axis.

The terrestrial conventional system adopted by IAG is called the International Terrestrial Reference System (ITRS). Its realization is the International Terrestrial Reference Frame (ITRF) maintained by the International Earth Rotation and Reference Systems Service's (IERS) ITRS Center. This is a set of points along with their three-dimensional Cartesian coordinates, respective point velocities, and a reference epoch. It is important to keep in mind that, at the present level of observational precision, the Earth (in particular the Earth's crust) is constantly in motion and deforming. A purely static description may be sufficient for some mapping or GIS applications. However, in view of the reference frames, kinematic modeling of points is mandatory for positioning and GNSS applications. Coordinates change by 1–3 cm/yr on average. Some regions, such as the Pacific and South East Asia, however, exhibit locations with velocities of up to 24 cm/yr. These motions can, for the majority of the solid surface of the Earth, be described quite well by plate motion models [14] – see Sect. 30.8.2 for a visualization – so long as one stays away from the plate boundaries. Within the deformation zones at the plate boundaries, irregular and large motion rates have to be expected, and are now modeled in the latest release of the ITRF, ITRF2014 [13].

The current procedure to compute the realization is to combine the observations from space geodesy techniques—VLBI, lunar and satellite laser ranging (LLR, SLR), GNSS, and Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS)—in a least-squares adjustment. The combination method makes use of local ties in collocation sites where two or more geodetic systems are being operated. ITRF solutions are published at intervals that depend on the number of new observations and the expected changes in the realization. The numbers (*yy* or *yyyy*) following the designation ITRF specify the last year of data that were used in the formation of the frame. Hence, ITRF98 designates the frame of station positions and velocities constructed using all of the IERS data available until 1998. The latest issue is ITRF2014 [13], with 975 sites. The coordinates, velocities, and full covariance matrix of the least-squares adjustment used for their estimation, along with the related uncertainties, can be obtained from the ITRF Product Center [7]. Actual positions of points at the time of observation or any other specific epoch in time t can be computed via

$$\boldsymbol{x}(t) = \boldsymbol{x}_0 + v(t - t_0) + \sum [\Delta \boldsymbol{x}_i(t)], \tag{8.3}$$

where $\boldsymbol{x}(t)$ is the vector of coordinates at epoch t, $\boldsymbol{x}_0$ is the vector of coordinates and velocities at the reference epoch t_0, $\boldsymbol{v}$ is the vector of velocities, and $\Delta \boldsymbol{x}_i(t)$ represents the site-specific time-dependent corrections. In ITRF2014, these site-specific corrections include discontinuities, annual and semiannual periodic motion components for stations operational for more than two years, and nonlinear motion components for 123 stations that were affected by earthquakes.

The reference epoch of an ITRF solution is specified to be the central epoch of the data span that was used in the computation. Presently, the reference epoch is January 1, 2010. Site-specific corrections are needed if the station position is to be determined with the highest accuracy—to the centimeter or subcentimeter level. In general, the station positions and velocities in the latest ITRF realization are accurate to the centimeter level. For national reference systems and networks or regional densifications, the coordinate solutions of the respective points are transformed into the ITRF by a suitable three-dimensional transformation (Sect. 8.8.3).

It is important to understand the concept of time-varying coordinates and the consequences for reference frames and coordinate determination. All computations, e. g., in GNSS positioning, should be performed with reference coordinates at the measurement epoch. Only the instantaneous coordinates correspond to measured coordinates or coordinate differences from actual geodetic observations. After that, the resulting coordinate sets may be transformed to any desired reference epoch as required, e. g., by some national reference system definitions. An example is the European Reference System [15], which is defined as a system with a static datum and a reference epoch at January 1, 1989. Therefore, after computing new coordinates in the actual ITRF at the observation epoch, the coordinates have to be rotated back into the reference epoch using the corresponding velocities [16]. For Central Europe for example, this corresponds to a coordinate change of about 80 cm if observed in 2018. For stations that do not have their own known velocity vector from repeated observations, the revised no-net-rotation Northwestern University velocity model (NNR-NUVEL-1A) for plate motion [14] or the plate motion model associated with the respective ITRF edition may be applied in stable regions. Countries that lie on an active fault, e. g., New Zealand, use a dynamic datum that is adapted to rapid and irregular changes in coordinates [17]. The typical procedure for computing the coordinates of a new site in ITRF by differential GNSS is:

1. Compute the baseline coordinate components from GNSS observations at epoch t.
2. Transform the known ITRF coordinates of the reference point to the actual values at the observation epoch t using its known ITRF velocities.
3. Compute the coordinates of the new station.
4. When the coordinates in a national reference frame at the respective reference epoch are needed, transform the resulting coordinates using, e. g., NUVEL plate motion velocities to the reference epoch.
5. As an alternative to step 4, for some national networks, transformation parameters are given that enable the trans-

Table 8.1 Parameters of the GRS80 and the WGS 84

Parameter	GRS80	WGS 84
Equatorial radius of the Earth (m)	6 378 137	6 378 137.0
Geocentric gravitational constant of the Earth, GM (m^3/s^2) (including the atmosphere)	$3\,986\,005 \times 10^8$	$(3\,986\,004.418 \pm 0.008) \times 10^8$
Reciprocal flattening of the reference ellipsoid	298.257222101	298.257223563
Angular velocity of the Earth, ω (rad/s)	$7\,292\,115 \times 10^{-11}$	$7\,292\,115 \times 10^{-11}$

formation of points in a certain region and pertaining to a particular network using empirically determined values that maintain the consistency of a particular reference frame realization. The backrotation of step 4 using model velocities may then be replaced with a three-dimensional seven-parameter coordinate transformation.

8.4.4 World Geodetic System (WGS 84)

The WGS 84 is a particular realization of the terrestrial reference system that is implicitly connected to the US GPS system. It is realized by the coordinates of the tracking station antennae of the GPS ground segment monitoring stations that are used to compute the broadcast ephemeris of the GPS satellites. In its predecessors and earlier versions before 1996 (WGS 72, for example), the WGS was determined exclusively from these observations. Since 1996, these tracking station coordinates have been computed from the ITRF coordinates and velocities, so WGS 84 is now identical to the ITRF [18]. Slight (centimeter-level) differences may occur, as the WGS may use an older version of the ITRF and the realization itself is less accurate than the primary International GNSS Service (IGS) stations used in the ITRF. Such differences do not matter in most navigation and positioning applications. By using the GPS broadcast ephemeris in GPS point positioning, the resulting coordinates will be in the ITRF at the observation epoch. When geodetic work of the utmost precision is needed, the methods described in the previous section have to be applied.

The WGS 84 definitions additionally include an associated WGS 84 ellipsoid to convert the Cartesian x, y, and z coordinates to ellipsoidal latitude, longitude, and height as well as a gravity field model to relate the ellipsoidal height to the geoid. The ellipsoidal flattening of the WGS 84 is slightly different from that of the GRS80 (Table 8.1) ellipsoid that should be used in geodesy; however, the difference for applications at or close to the Earth's surface is negligible. The gravity field model that is used to compute the geoid as a global vertical reference surface is the enhanced Earth Gravitational Model 1996 (EGM96), a spherical harmonic expansion complete to degree and order 360. Using this model, the height above mean sea level can be computed as $H = h - N$ to an accuracy of better than 1 m anywhere on the Earth.

8.4.5 Geodetic Reference System 1980

The Geodetic Reference System 1980 (GRS80) is the official geodetic reference system recommended by the International Association of Geodesy. It should be used in all geodetic work and in computations of the reference gravity field, both on the Earth's surface and in outer space, and is defined in [19]. Four parameters are used to uniquely define the best-fit ellipsoid and the reference gravity field (normal gravity and potential) of the Earth based on the theory of the geocentric equipotential ellipsoid. The defining conventional constants for GRS80 are given in Table 8.1.

Its origin and orientation are such that the minor axis of the reference ellipsoid, defined above, is parallel to the direction defined by the Conventional International Origin, and such that the primary meridian is parallel to the zero meridian of the BIH-adopted longitudes, which is coincident with the x, y, z Cartesian coordinate system constituted by the ITRF. The GRS80 should eventually replace all nongeocentric regional reference systems and geodetic datums. Together with the ITRF, it provides the basis for the computation of globally homogeneous coordinates, mappings, and reference gravity field parameters.

8.5 Height Systems and the Vertical Datum

8.5.1 Definition of Height in Geodesy

The term *height* is intrinsically problematic in geodesy, as its definition can be purely geometric or based on physics, i. e., on the potential of the Earth. There is no "best" height, as the use of a particular height definition depends on the application. Ellipsoidal heights are defined in a purely geometric way and are not suitable for technical purposes. They are defined as the distance from a point to the chosen reference ellipsoid along the ellipsoidal normal.

Heights based on the potential are measured not in meters but in potential differences. The difference in potential $W_B - W_A$ of two points is the work done in transporting a unit mass of 1 kg from A to B. This work is independent of the path taken from A to B. Potential differences are measured by a combination of geometric leveling and gravity measurements. This is known as geopotential leveling. The physical unit of potential difference is $m^2\,s^{-2}$. As users gen-

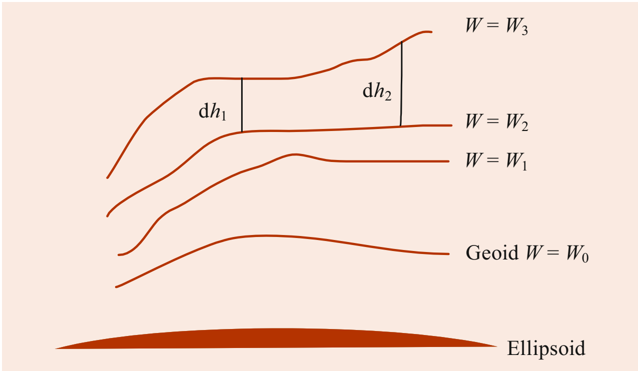


Fig. 8.9 Depictions of the ellipsoid for which the geometric surface is identical to the equipotential surface of the normal potential U_0, the geoid, and various nonparallel equipotential surfaces W_i

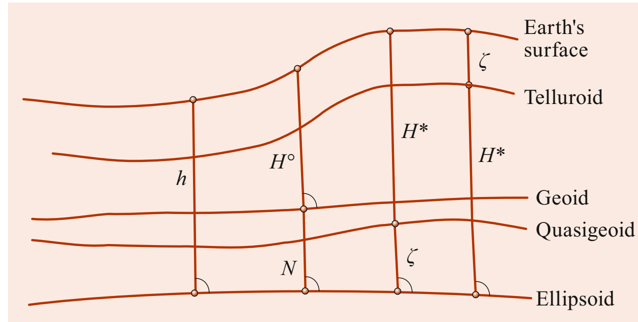


Fig. 8.10 Height types (ellipsoidal height h, orthometric height H^{O}, normal height H^*) and the corresponding reference surfaces, along with their respective distances to the ellipsoid, the geoid height N, and the height anomaly ζ

erally prefer a geometric value in meters, a suitable definition of a metric height system has to be agreed on. It should fulfill three requirements [20]:

- The height of a point should be unambiguous and independent of the way it was measured
- The height should ideally be free of any hypothesis
- Corrections made to height differences measured via geometric leveling to obtain the system adopted should be small so that neglecting them does not have a major effect.

As seen in Fig. 8.9, equipotential surfaces are not parallel, i. e., points with the same potential do not have the same ellipsoidal height, and the potential-related height may vary both with horizontal and/or vertical position. In order to eliminate these variations that result in a path dependency of geometric leveling, gravity values along the leveling lines have to be known. For a potential increment of $\mathrm{d}W$, a gravitational acceleration of g, a height increment of $\mathrm{d}h$, and a leveled height difference of δh, the basic relation to compute the potential difference is $\mathrm{d}W = -g\mathrm{d}h$. Integration along the leveling line gives the potential difference $W_{\mathrm{A}} - W_{\mathrm{B}}$ between a starting point A and an end point B as

$$W_{\mathrm{A}} - W_{\mathrm{B}} = \int_{\mathrm{A}}^{\mathrm{B}} g\mathrm{d}h \approx \sum_{\mathrm{A}}^{\mathrm{B}} g\delta h. \quad (8.4)$$

When defining a height system, a zero level has to be established. This is accomplished by using the geoid, the potential of which is labeled W_0 in the corresponding relations. The potential difference from this value W_0 is called the geopotential number (C_{A}), where

$$C_{\mathrm{A}} = W_0 - W_{\mathrm{A}} = \int_{\mathrm{P}_0}^{\mathrm{A}} g\mathrm{d}h \approx \sum_{\mathrm{P}_0}^{\mathrm{A}} g\delta h. \quad (8.5)$$

The geopotential number can be formally converted to a metric quantity by dividing it by a gravity value. The value used in Europe is the gravity value of the reference ellipsoid at 45° latitude, called normal gravity (γ_{45}). The result is the dynamic height,

$$H_{\mathrm{A}} = \frac{C_{\mathrm{A}}}{\gamma_{45}}. \quad (8.6)$$

Dynamic heights are strict, as no water can flow between two points with identical heights and there are no hypotheses involved. Their main drawback is the fact that they require large corrections to the leveled height differences, and they cannot be combined with GNSS heights as they do not have a defined zero level and no geometric interpretation.

8.5.2 Orthometric Height (H^{O})

The orthometric height, H^{O}, is the length of the slightly curved plumb line from the geoid to the Earth surface (Fig. 8.10). The orthometric height usually reflects local variations in gravity as well as changes in topography. A fictitious geopotential leveling along the plumb line from the geoid to a point A will give the geopotential number C_{A}. Because C_{A} is independent of the leveling path, the same value results from a geopotential leveling along the Earth's surface. The relation between $H_{\mathrm{A}}^{\mathrm{O}}$ and C_{A} is

$$C_{\mathrm{A}} = W_0 - W_{\mathrm{A}} = \int_{\mathrm{A}_0}^{\mathrm{A}} g\mathrm{d}h = h_{\mathrm{A}} \frac{1}{h_{\mathrm{A}}} \int_{\mathrm{A}_0}^{\mathrm{A}} g\mathrm{d}h = \overline{g}_{\mathrm{A}} H_{\mathrm{A}}^{\mathrm{O}}, \quad (8.7)$$

$$H_{\mathrm{A}}^{\mathrm{O}} = \frac{C_{\mathrm{A}}}{\overline{g}_{\mathrm{A}}^*}, \quad (8.8)$$

where $\overline{g}_{\mathrm{A}}^*$ is the integral mean of the gravity along the plumb line, which has to be computed from gravity values measured at the Earth's surface. This is where hypothetical assumptions about the density enter the calculations. As a consequence, it is rarely possible to get orthometric heights with millimeter accuracy, and even centimeter accuracy may not be attainable in mountainous areas. If we consider the three previously defined requirements for a good height system, the first is fulfilled, as is the third to a large extent, but the second is not. One drawback is that water can flow between two points with equal H^{O}. However, the most important advantage of H^{O} values is that they can be combined with ellipsoidal heights: The difference in ellipsoidal height equals the difference in orthometric height plus the difference in geoidal undulation, i. e.,

$$(h_{\mathrm{B}} - h_{\mathrm{A}}) = (H_{\mathrm{B}} - H_{\mathrm{A}}) + (N_{\mathrm{B}} - N_{\mathrm{A}}). \tag{8.9}$$

8.5.3 Normal Heights (H^*)

Normal heights are the most advanced concept. They are related to the geodetic theory of the Russian geodesist Molodenskij. In his theory, the surface of the Earth is mapped point by point onto another surface. Each point A on the surface of the Earth receives a partner point Q on the same ellipsoid normal above or below A. The ellipsoidal height of Q depends on values that can be calculated without hypotheses, which is what makes Molodenskij's theory so attractive.

The normal height H^* of point A is defined as the ellipsoidal height of the partner point Q. The height of Q is calculated from potential differences. The basics are formulated according to

$$U_0 - U_{\mathrm{Q}} = W_0 - W_{\mathrm{A}} = C_{\mathrm{A}}, \tag{8.10}$$

where U_0, the mean Earth ellipsoid, generates an associated theoretical gravity field. Its potential, labeled U_0, is by definition equal to the potential W_0 of the geoid (Fig. 8.9). C_{A}, the geopotential number, can be determined from (8.5).

Thus, U_{Q} is computable without hypotheses as

$$U_{\mathrm{Q}} = U_0 - C_{\mathrm{A}} = W_{\mathrm{A}} - W_0 + U_0.$$

The set of all points Q represents a surface that is similar but not identical to the Earth's surface. This surface is called the *telluroid*, and *is not an equipotential surface*. The distance between the Earth's surface and the telluroid ζ is termed the *height anomaly*.

Similar to the formula for the orthometric height (8.7), the relation for the normal height is

$$C_{\mathrm{A}} = U_0 - U_{\mathrm{Q}} = \int_0^{H^*} \gamma \, \mathrm{d}H_{\mathrm{A}}^* = \overline{\gamma}_{\mathrm{Q}} H_{\mathrm{A}}^*.$$

This leads to the normal height of point A,

$$H_{\mathrm{A}}^* = \frac{C_{\mathrm{A}}}{\overline{\gamma}_{\mathrm{Q}}}, \tag{8.11}$$

where $\overline{\gamma}_{\mathrm{Q}}$ is the integral mean of the theoretical gravity from the ellipsoid to point Q, which can be computed as shown from the normal gravity formula once the iterative determination of point Q has been accomplished.

The quasigeoid is another surface commonly used in geodetic science. The distance between a point A on the Earth's surface and the quasigeoid is exactly the normal height of the point H_{A}. The pair of the geoid height N and the orthometric height are equivalent to the pair of the quasigeoid height and the normal height.

The geometric interpretation can be derived from Fig. 8.10 as

$$(H_{\mathrm{B}}^{\mathrm{O}} - H_{\mathrm{A}}^{\mathrm{O}}) + (N_{\mathrm{B}} - N_{\mathrm{A}}) = (H_{\mathrm{B}}^* - H_{\mathrm{A}}^*) + (\zeta_{\mathrm{B}} - \zeta_{\mathrm{A}}). \tag{8.12}$$

In analogy to the telluroid, the quasigeoid is not an equipotential surface. It coincides by definition with the geoid on the oceans. On the continents, it runs slightly above the geoid. The difference between N and ζ depends on the geology and the topography itself; in mountainous regions, this difference may amount to 40–50 cm.

8

The importance of orthometric and normal heights lies in the fact that, as soon as the detailed geoid (or quasigeoid) is known, costly leveling operations can be replaced with GNSS observations. Computation of the necessary height anomalies as well as the geoid undulations is the task of physical geodesy, and is a very ambitious problem. Regional models allow the computation of the geoid or quasigeoid height to centimeter precision in a limited area. GNSS leveling—the determination of physical heights by combining ellipsoidal heights from GNSS and geoid or quasigeoid heights—is therefore about to replace traditional leveling at the national and regional scales [21]. However, global geopotential models are not yet sufficiently accurate to allow the absolute N or ζ to be computed to better than the decimeter level (Sect. 8.6).

A comparison of illustrative values of the four types of heights in use is given in Table 8.2. The data are taken from points on the Austrian first-order leveling net ranging from the lowlands to the Alps.

The particular height system used depends on the country; however, for Europe, the use of normal heights is recommended by the EUREF Subcommission of the IAG for Europe [23]. Normal heights based on an adjustment of geopotential numbers are recommended as a standard. Future height systems intended for precise applications in monitoring and geodynamics will have to take height variations over time into consideration.

Vertical Datum

The zero surface to which elevations or heights are referred to is called the vertical datum. Traditionally, it is associated with mean sea level. The mean sea level (MSL) reference

Table 8.2 Comparison of different height types (Source: [22])

Geopotential number C (kGal m)	Dynamic height (m)	Orthometric height (m)	Normal height (m)	$(\zeta - N)$ (m)	Geoidal undulation N (m)	Ellipsoidal height (m)
140.0704	142.839	142.801	142.800	+0.001	+0.60	143.40
694.0876	707.805	707.810	707.721	+0.089	+1.18	708.99
1090.1256	1111.670	1111.797	1111.645	+0.152	+2.35	1114.15
1420.8730	1448.954	1449.340	1449.037	+0.303	+2.77	1452.11

is realized by taking continuous measurements at tide gage stations. The average reading over a sufficiently long interval is the MSL, which is then used as zero elevation for a local or regional area. The MSL is an approximation to the geoid. Local differences arise because the sea level includes position- and time-dependent components due to currents, winds, tides, and salinity, among others, which cause deviations from an equipotential surface of the Earth's gravity field. This sea surface topography part is typically in the decimeter range; globally it ranges between -1.5 and $+1.5$ m. Therefore, national height systems that refer to different tide gages may have offsets in the meter range.

A unified global vertical datum is needed for georeferencing and for global monitoring of the Earth, as this will lead to greater accuracy when connecting national and continental datums. It will also improve geoid computations, as systematic regional biases in gravity anomaly databases are removed by referring gravity anomalies to one unique geopotential surface. However, due to a whole range of fundamental geodetic questions such as the choice of the reference potential of the mean Earth ellipsoid, the value of the potential W_0 for the geoid and its time dependency, plus many more, there is no clear definition at the moment. Presently, the EGM gravity models are the de facto standard when used with ITRF-derived ellipsoidal heights from GNSS to compute H^{O} or H^*. Both height systems (i. e., that derived from tide gages and leveling and that derived from GNSS leveling and geopotential models) are in use.

Height datums and height values, like all quantities that stem from observations, can be inconsistent for several reasons. In principle, the geometric heights h may change by δh due to changes in the reference ellipsoid. The physical height H may change by δH due to changes in gravity, leveling, or reference potential at the tide gage or the selected zero point. Geoid heights N may change by δN due to changes in the reference potential W_0 or geoid redefinitions. These effects generally appear as a near-constant bias in a given area, which may be expressed by

$$h + \delta h = (H + \delta H) + (N + \delta N). \qquad (8.13)$$

The necessity to combine geometric coordinates and a potential-based height led to the introduction of compound coordinate reference system descriptions. A compound coordinate reference system describes a position via two independent coordinate reference systems, e. g., an ITRF-based geometry component and a height system with a global geopotential model and an associated vertical datum. Presently, the IAG is in the process of defining and establishing such an International Height Reference System (IHRS). This will allow the realization of a high-precision global physical reference frame at the centimeter level that will provide a reliable vertical datum and allow the detection of changes in the Earth's gravity field related to sea level changes and other physical effects [24].

8.6 Geopotential Models and the Geoid

This section gives a brief description of geopotential and geoid models in the framework of georeferencing and GIS applications. These models are used to compute physical heights above the geoid in combination with GNSS leveling, the realization and unification of the vertical reference system, and the transformation of local geodetic observations to a global reference frame. They have to be consistent with the geometric terrestrial reference system, ITRF. In practice, a global representation of the geopotential of the Earth obtained by an expansion in spherical harmonics [20] is used. Until 2008, the IAG-recommended global gravity field was the Earth Gravity Model 96 (EGM96), computed by the US National Geospatial-Intelligence Agency NGA (formerly known as the Defense Mapping Agency (DMA) or National Imagery and Mapping Agency (NIMA)) in cooperation with the Ohio State University [25]. These data comprise a set of fully normalized Earth-gravity (geopotential) coefficients, complete to degree (n) and order (m) 360, corresponding to resolution of the gravity field's features at the 100 km scale. The spherical harmonic coefficients are used in Clenshaw summation numerical algorithms or fast Fourier transformation algorithms to compute all quantities of interest in geodesy: the geopotential W, point gravity anomalies, point geoid heights, the point N–S or E–W component of the deflection of the vertical, the point total deflection of the vertical, the point radial component of the gravity disturbance vector, and the point N–S or E–W component of the gravity disturbance vector. Detailed descriptions of the background to this topic and application formulas are published in [26]. In 2009, EGM96 was replaced by EGM2008 [25]. This set of coefficients is based on a combination of terrestrial gravity observations and

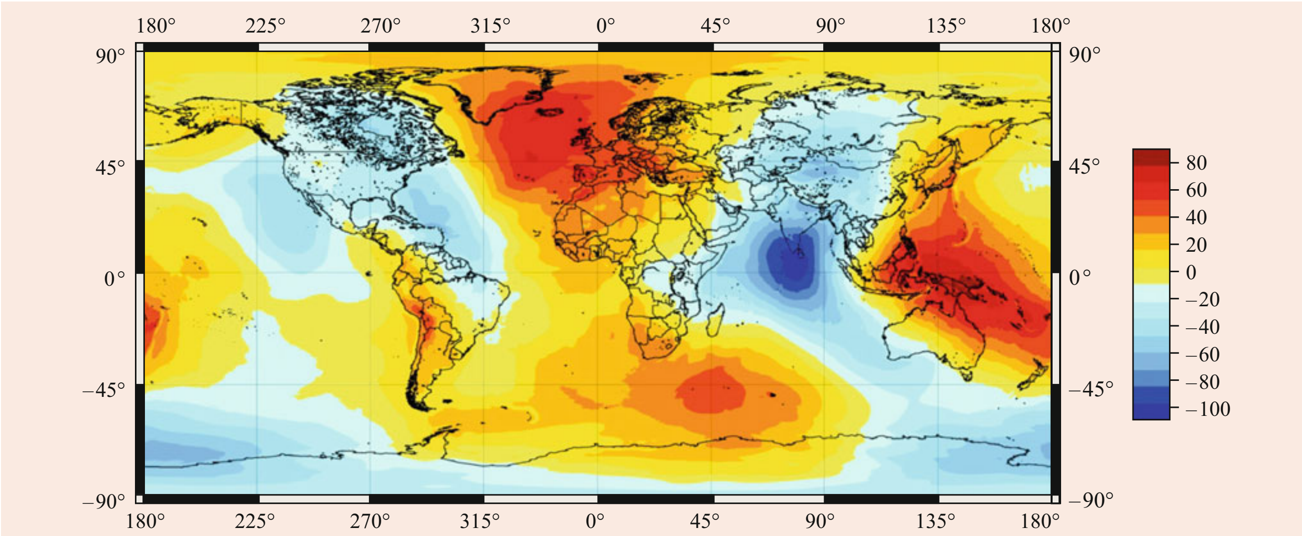


Fig. 8.11 EGM2008 geoid heights (m)

data from satellite gravity observations. The latter were obtained from satellite laser-ranging missions, satellite altimetry missions over the oceans, and, more recently, the Challenging Minisatellite Payload (CHAMP), Gravity Recovery and Climate Experiment (GRACE), and Gravity Field and Steady-State Ocean Circulation (GOCE) dedicated satellite gravity missions [27]. The latter missions have led to significant improvements in the modeling of long-wavelength gravity signals. When used in combination with terrestrial gravity data of good quality (± 1 mGal, where 1 mGal $= 10^{-5}\,\mathrm{m/s^2}$) and coverage, significantly improved continental-scale geoid and quasigeoid models are obtained. The accuracy and resolution of EGM2008 are significantly enhanced compared to its predecessor (Fig. 8.11); the spatial resolution is now about 15 km or degree and order 2159 in the frequency domain. The accuracy of geoid or quasigeoid height computations is improved from about 0.3 m with EGM96 to 0.13 m with EGM2008, as shown by comparisons with selected GNSS leveling results. The accuracy may, however, vary with the region of the world considered, as it depends on the terrestrial input data available for the EGM2008 computation, and reaches 0.5–1 m in extreme cases. The average deviation of the global geoid undulations with respect to the WGS 84 ellipsoid is about 30.5 m, with minima and maxima of -107 and 85 m, respectively. These values illustrate the deviation of the geoid from the ellipsoid and the error that may arise if ellipsoidal heights and physical heights above mean sea level are not distinguished.

For GIS applications, the gridded data set of EGM2008 and a particular version for use with WGS 84 are available from the NGA website, just as for previous models [28]. These have resolutions of 1, 2.5, and 5 arcmin. Software for the synthesis of harmonic coefficients or for interpolating the grids to particular positions on the Earth is provided by NGA as well. The International Center for Global Earth Models (ICGEM) provides a web-based access to all models, including the latest satellite-only and combination models, tools for calculation, and background information [29].

8.7 Time Systems

Time systems play a fundamental role when dealing with space geodesy and advanced GNSS data analysis. The typical example is the transformation between the space-fixed ICRS and the Earth-fixed ITRF implicitly included in GNSS positioning. Sensor systems and sensor fusion depend on precise timing and time synchronization. Moreover, time is the fundamental quantity that forms the basis of almost all modern geodetic observation techniques used in geodesy and GIS [30]. It is measurable to the level of one part in 10^{15}. Recent developments in optical clocks have led to an improvement in accuracy of two orders of magnitude, which will have a huge impact on GNSSs and positioning in the next 5–10 years [31]. This section gives a brief introduction to the timescales used in GIS and geodesy. Basically, we have to differentiate between the unit of time (e. g., the interval of 1 s) and the epoch, i. e., a particular instance of an event in time. Each time system may have its own unit and its own zero epoch.

There are four basic time systems that are in use:

1. *Solar time* is based on the daily path of a fictitious sun that moves with constant velocity along the equator. This is the basis of Universal Time (UT). One second is 1/86 400 of a solar day. Because of the changes in rotational speed of the Earth and the slowing of the Earth's rotation, it is not constant.
2. *Sidereal time* is based on the rotational speed of the Earth. The unit is the period of the Earth's rotation with respect

to a point nearly fixed with respect to the stars. One sidereal day is 4 min shorter than the mean solar day due to the revolution of the Earth around the sun.

3. *Atomic time* (TAI) has a fundamental interval of one Système International (SI) second. It is defined as the duration of 9 192 631 770 cycles of radiation corresponding to the transition between two hyperfine levels of the ground state of cesium-133 (^{133}Cs). The SI day is defined as 86 400 s, and the Julian century as 36 525 days. TAI is the International Atomic Time scale, a statistical timescale based on a large number of atomic clocks. The origin was established as January 1, 1958. At midnight on January 1, 1958, Universal Time and sidereal time effectively ceased to function as time systems.
4. *Dynamic time* is based on the equations of motion of the celestial bodies in the solar system. The theory of general relativity implies that we have to select an adequate inertial reference frame. For events at or close to the Earth, it is suitable to use Terrestrial Time (TT), which is the successor to Terrestrial Dynamical Time and represents a uniform timescale for motion within the Earth's gravity field. By definition, it has the same rate as an atomic clock on the geoid. TAI is related to the definition of TDT by the definition
$$\mathrm{TDT} = \mathrm{TAI} + 32.184\,\mathrm{s}. \tag{8.14}$$

8.7.1 Timescales and GNSS Times

TAI is a continuous timescale and so does not remain synchronized with the mean solar day (UT1), since the Earth's rotation is slowing by an average of about 1 s/yr. This problem is taken care of by defining Coordinated Universal Time (UTC), which runs at the same rate as TAI but is incremented by leap seconds periodically. Leap seconds are introduced by the IERS so that UTC does not vary from UT1 by more than 0.9 s. As of January 2021, the difference amounts to

$$\mathrm{UTC} - \mathrm{TAI} = -37. \tag{8.15}$$

GPS time is derived from TAI. The time signals broadcast by the GPS satellites are synchronized with the atomic clock at the GPS Master Control Station in Colorado, USA. Zero in Global Positioning System time (GPST) was set to 0 h UTC on January 6, 1980. It is not incremented by UTC leap seconds. Therefore, there is an integer-second offset of 19 s between GPST and TAI such that

$$\mathrm{GPST} + 19\,\mathrm{s} = \mathrm{TAI}. \tag{8.16}$$

As of 2021, there have been a total of 17 leap seconds since January 6, 1980. Therefore, currently,

$$\mathrm{GPST} + \mathrm{UTC} = 17\,\mathrm{s}. \tag{8.17}$$

For precise applications, this offset between UTC and GPST has to be adequately considered by specifying the time system used. GPS time is primarily counted in GPS week numbers and seconds of a week. Since January 6, 1980, each week has been designated its own number. For example, February 4, 2009 is day of year 35 in GPS week 1517. A given epoch within the week is identified using the concept of *seconds of week*. This number counts from midnight between Saturday and Sunday, the beginning of the GPS week. Furthermore, for convenience, the individual days of the week are numbered as follows: Sunday, 1; Monday, 2; Tuesday, 3; Wednesday, 4; Thursday, 5; Friday, 6; and Saturday, 7. Professional GPS software uses day of week and seconds of day for numerical reasons.

Other GNSSs, such as the Russian Global Navigation Satellite System (GLONASS) or the European Galileo system, maintain their own time systems. However, like GPS, they are realizations of UTC and are steered to be within 1 μs of UTC, modulo one second. GNSS times, except GLONASS, are not adjusted for leap seconds. Their mutual offsets are broadcast to users to allow interoperability and seamless use of all GNSSs.

A continuous time count often used in astronomy, geodesy, and GIS is the Julian date (JD). It describes the number of days and the fraction of a day after a zero epoch that is sufficiently in the past to precede the historical record. This zero epoch was chosen to be 12 h UT on January 1, 4713 BC. The JD of the standard epoch of UT is called J2000.0, where

$$\begin{aligned} \mathrm{J2000.0} &= \mathrm{JD}\,2\,451\,545{:}\,0 = 2000\ \mathrm{January}\ 1.5^{\mathrm{d}}\ \mathrm{UT} \\ &= \mathrm{January\ 1st,\ 12h\,UT}. \end{aligned} \tag{8.18}$$

The JD is a large number, so often it is replaced by the modified Julian date (MJD), where

$$\mathrm{MJD} = \mathrm{JD} - 2\,400\,000.5. \tag{8.19}$$

Hence J2000.0 = MJD 51544.5. MJD, in contrast to JD, starts at midnight.

8.8 Conversions, Transformations, and Projections

There are two basic kinds of coordinate operations: coordinate transformation and coordinate conversion.

Transformations are basic operations in geodesy. They include coordinate transformations between different types of coordinate systems (Fig. 8.12) as well as linear, affine, and projective transformations. Typically, they comprise translation, rotation, and a change in scale. Formulas are available for Cartesian or ellipsoidal coordinates; the relevant set of

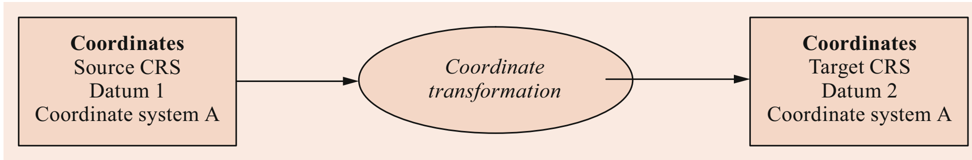


Fig. 8.12 Schema for coordinate transformation (source: [10])

formulas are known as the Helmert transformation and the Molodenskij formulas, respectively. The latter also include terms to account for the change in ellipsoid parameters, i. e., the dimensions of the ellipsoid. They were used in geodetic datum transformations, e. g., to relate nongeocentric geodetic systems with ellipsoidal coordinates to the ITRF and the GRS80. Today, GNSS-based coordinates are Cartesian and are given in the ITRS or WGS 84; therefore, the Helmert transformation is appropriate in most cases. The transformation parameters are, in general, derived empirically in a least-squares estimation using a set of identical points known in both systems. The choice, allocation, number, and quality of the coordinates of those points extensively affect the results and the accuracy.

For a three-dimensional CRS, the seven-parameter Helmert transformation is generally used for coordinate transformations (see (8.36)–(8.38)). For two-dimensional and geotopographic data, a grid-based transformation is also applicable. The ellipsoidal coordinates of the identical points are computed first and stored in a regular grid. The shifts needed for the transformation of other or new coordinates are then computed by bilinear interpolation inside the grid meshes.

The change from one coordinate system to another based on the same datum is accomplished by a coordinate conversion. In this case, mathematical rules are specified. Generally, these conversions are unambiguous and can be realized with high accuracy. Examples are map projections and conversions between Cartesian and ellipsoidal coordinates.

The change of coordinates from one CRS to another may result from a series of operations consisting of one or several transformations and conversions by concatenated operations.

8.8.1 Conversion Between Ellipsoidal and Cartesian Coordinates

A conversion between Cartesian coordinates (e. g., those obtained from the ITRF) and ellipsoidal coordinates that are used to describe the location of a point on or above the ellipsoid is often required in positioning and GISs. An ellipse or ellipsoid is defined by two parameters: its semimajor axis a and its semiminor axis b. The equation of an ellipsoid of revolution in a Cartesian system with its origin at the center and the z-axis along the minor axis is given by

$$\frac{x^2+y^2}{a^2}+\frac{z^2}{b^2}=1. \tag{8.20}$$

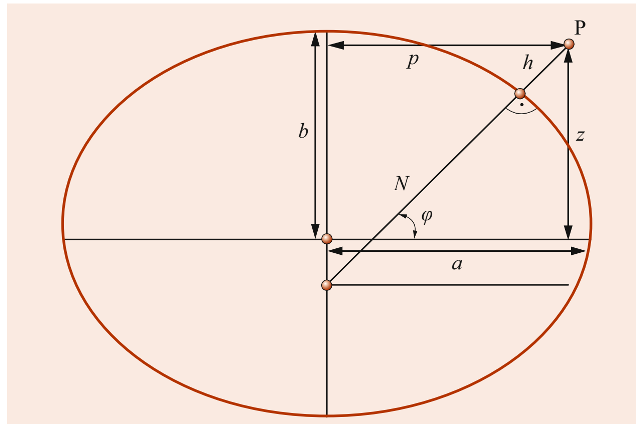


Fig. 8.13 Geometry of the ellipsoid

A number of frequently used quantities that describe the geometry of an ellipsoid of rotation can be derived by examining a meridian curve of the ellipsoid (Fig. 8.13): the flattening (f), the first eccentricity (e), the second eccentricity (e'), and the radius of polar curvature (c), where

$$f=\frac{a-b}{a}; \qquad e^2=\frac{a^2-b^2}{a^2};$$
$$e'^2=\frac{a^2-b^2}{b^2}; \qquad c=\frac{a^2}{b}. \tag{8.21}$$

With the aid of the auxiliary quantities

$$V=\sqrt{1+e'^2\cos^2\varphi},$$
$$N=\frac{c}{V},$$
$$M=\frac{c}{V^3}, \tag{8.22}$$

where M and N are the radii of curvature of the ellipsoid in the direction of the meridian and orthogonal to it, the transformation of the ellipsoidal coordinates to the Cartesian system is given by

$$x=\left(\frac{c}{V}+H\right)\cos\varphi\cos\lambda,$$
$$y=\left(\frac{c}{V}+H\right)\cos\varphi\sin\lambda,$$
$$z=\left[\frac{c}{V}(1-e^2)+h\right]\sin\varphi,$$

or

$$x=(N+h)\cos\varphi\cos\lambda,$$
$$y=(N+h)\cos\varphi\sin\lambda,$$
$$z=[N(1-e^2)+h]\sin\varphi. \tag{8.23}$$

The inverse transformation is not as straightforward; λ follows from

$$\lambda = \arctan\left(\frac{y}{x}\right). \tag{8.24}$$

Solving (8.23) for φ and h is theoretically possible but very complicated. In practice, this is done using an iterative procedure starting with $h = 0$, or using an approximate solution that is nevertheless rather accurate. The auxiliary terms θ and p are defined by

$$\theta = \arctan\left(\frac{az}{bp}\right), p = \sqrt{x^2 + y^2}. \tag{8.25}$$

The latitude φ and height h are obtained from

$$\varphi = \arctan\left(\frac{z + e'^2 b \sin^3\theta}{p - e^2 a \cos^3\theta}\right), \tag{8.26}$$

$$h = \frac{p}{\cos\varphi} - \frac{c}{V}. \tag{8.27}$$

Care has to be taken to retain millimeter accuracy; the latitude and longitude have to be given to an accuracy of $0.0001'$ or 3×10^{-8}. Two normals to the ellipsoid that subtend the small angle of $1'$ intersect the ellipsoid at two points 30 m apart.

For many computations in a limited area, the ellipsoid can be substituted by an osculating sphere situated tangent to the ellipsoid at a central point in the region. The radius R of this sphere is equal to the geometric mean of the principal radii M and N of the ellipsoid M and N, i.e.,

$$R = \frac{c}{V^2}. \tag{8.28}$$

V has to be calculated for the latitude of the tangent point.

Two different types of latitude at the ellipsoid should be distinguished: the reduced latitude β and the geocentric latitude γ. They are used in connection with the gravity models of the geopotential (Sect. 8.6). As seen in Fig. 8.14, the reduced latitude β is computed from a circle of radius a with the vertical through the point at the ellipsoid from the components p and a, whereas the geocentric latitude γ is computed from z and p.

$$\beta = \arccos\left(\frac{p}{a}\right)$$
$$\gamma = \arctan\left(\frac{z}{p}\right) \tag{8.29}$$

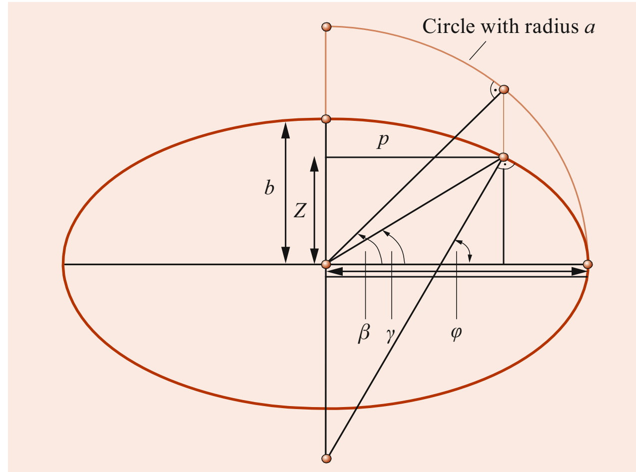


Fig. 8.14 Geodetic latitude φ, reduced latitude β, and geocentric latitude γ

8.8.2 Local Geodetic Systems

To handle terrestrial data, local observation systems such as that shown in Fig. 8.15 have to be related to global coordinates. The origin of a local system is topocentric, i.e., it is located at some point P on the surface of the Earth with orthogonal axes u (in the direction of geodetic north), v (to the east), and w (parallel to the local normal on the ellipsoid). This system is also termed the *horizon system*. A local observation at point P to point Q can be made by observing the distance s, the azimuth α, and the zenith distance ζ. Note that actual observations have to be corrected for the deflection of the vertical, i.e., the difference between natural coordinates and ellipsoidal coordinates. The vector $\boldsymbol{s}$ can be computed as

$$\boldsymbol{s} = \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} \cos\alpha\sin\zeta \\ \sin\alpha\sin\zeta \\ \cos\zeta \end{pmatrix}. \tag{8.30}$$

Now the vector $\boldsymbol{s}$ has to be transformed into the difference vector $\boldsymbol{S}$ of the position vectors of points P and Q in the global CRS and its associated ellipsoid. As shown in Fig. 8.15, the two Cartesian systems have different orientations, so the conversion will include two rotations for latitude and longitude and one mirroring to convert the left-handed local system into the right-handed geocentric system. The transformation then reads

$$\begin{aligned}\boldsymbol{S} &= (\Delta x, \Delta y, \Delta z)^\top \\ &= \boldsymbol{R}_w(180° - \lambda)\boldsymbol{R}_v(90° - \varphi)\boldsymbol{R}_v \cdot s = R \cdot s,\end{aligned} \tag{8.31}$$

with

$$\boldsymbol{R} = \begin{pmatrix} -\sin\varphi\cos\lambda & -\sin\lambda & \cos\varphi\cos\lambda \\ -\sin\varphi\sin\lambda & \cos\lambda & \cos\varphi\sin\lambda \\ \cos\varphi & 0 & \sin\varphi \end{pmatrix}. \tag{8.32}$$

For the inverse transformation, with the difference vector $\boldsymbol{S} = (\Delta x, \Delta y, \Delta z)^\top$, we get

$$s = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}, \tag{8.33}$$

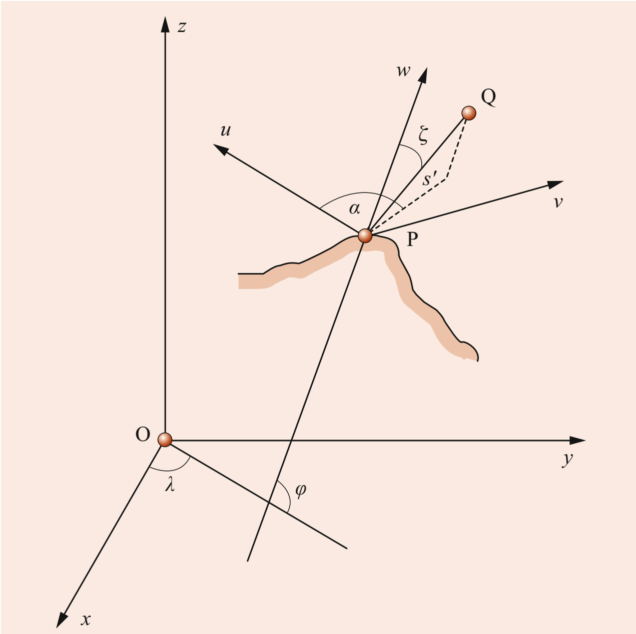


Fig. 8.15 Local observations in the topocentric system and their relation to the global CRS

$$\alpha = \arctan\left[(-\sin\lambda \cdot \Delta x + \cos\lambda \cdot \Delta y)/\right.$$
$$\left.(-\sin\varphi\cos\lambda \cdot \Delta x - \sin\varphi\sin\lambda \cdot \Delta y + \cos\varphi \cdot \Delta z)\right], \quad (8.34)$$

$$\zeta = \arccos\left[\frac{1}{s}(\cos\varphi\cos\lambda \cdot \Delta x + \cos\varphi\sin\lambda \cdot \Delta y + \sin\varphi \cdot \Delta z)\right]. \quad (8.35)$$

8.8.3 Coordinate Transformation and Terrestrial Frame Transformation

Modern reference frames based on satellite geodesy are orthogonal and homogeneous. The standard relation for transforming between two reference systems is a Euclidian similarity transformation with seven parameters: three translation components, one scale factor, and three rotation angles. At the level of accuracy of ITRF2014, the transformation parameters may be partly time dependent as different definitions are used for the initial adjustment of a particular frame. Therefore, the velocities must also be transformed using the time derivative of the seven parameters. Detailed transformation formulas and the most recent estimates for the transformation parameters are available at the ITRF website [8].

The transformation of coordinate vector $\boldsymbol{x}$ expressed in a reference system S into a coordinate vector $\boldsymbol{x}'$ expressed in a reference system S' is computed via (8.36), which consists

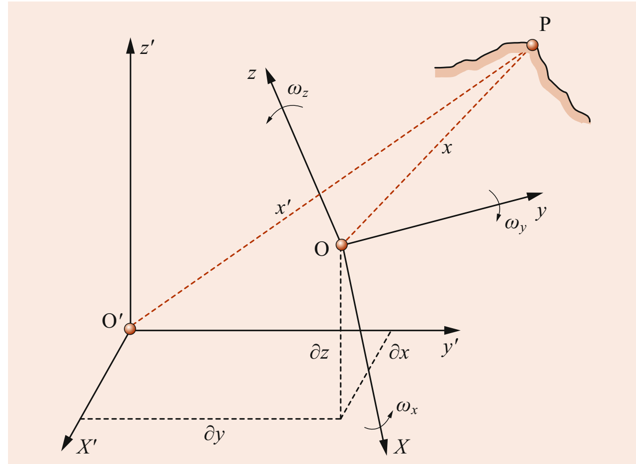


Fig. 8.16 Three-dimensional spatial coordinate transformation

of a translation for the shift in origin, a rotation, and a scale change along all three axes (Fig. 8.16), i. e.,

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \partial x \\ \partial y \\ \partial z \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (8.36)$$

The rotation matrix also includes a scale change. In a coordinate transformation, this affine mapping is constrained by implying conformity through the use of only one scale factor m. This leads to the standard seven-parameter similarity transformation widely used in geodesy,

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \partial x \\ \partial y \\ \partial z \end{pmatrix} + (1+m) \cdot R \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad (8.37)$$

$$\boldsymbol{R} = \begin{pmatrix} 1 & \omega_z & -\omega_y \\ -\omega_z & 1 & \omega_x \\ \omega_y & -\omega_x & 1 \end{pmatrix}. \quad (8.38)$$

This transformation is known as the spatial or 3-D seven-parameter Helmert transformation. In general, the rotation angles ω_i around the three coordinate axes are quite small, so the simple form given in (8.38) can be used. The transformation parameters are determined by using at least three identical points and solving the equation for the seven parameters. In general, this is accomplished by least-squares adjustment.

For the transformation of sets of ellipsoidal coordinates to a new reference ellipsoid, an alternative to the computation of the new latitude, longitude, and ellipsoid height involving the concatenation of three operations (geographic to geocentric, geocentric to geocentric, geocentric to geographic) is to use formulas derived by *Molodenskij* [32]. These directly relate the changes in geographical coordinate offsets caused by using the transformation parameters for the origin to the changes in ellipsoid parameters. Their short forms are

$$\begin{aligned}
\varphi' &= \varphi + \mathrm{d}\varphi, \\
\lambda' &= \lambda + \mathrm{d}\lambda, \\
h' &= h + \mathrm{d}h, \\
\mathrm{d}\varphi'' &= [-\mathrm{d}X \sin\varphi \cos\lambda - \mathrm{d}Y \sin\varphi \sin\lambda + \mathrm{d}Z \cos\varphi \\
&\quad + (a\mathrm{d}f + f\mathrm{d}a)\sin 2\varphi]/(M \sin 1''), \\
\mathrm{d}\lambda'' &= \frac{-\mathrm{d}X \sin\varphi + \mathrm{d}Y \cos\lambda}{N \cos\varphi \sin 1''}, \\
\mathrm{d}h &= \mathrm{d}X \cos\varphi \cos\lambda + \mathrm{d}Y \cos\varphi \sin\lambda + \mathrm{d}Z \sin\varphi \\
&\quad + (a\mathrm{d}f + f\mathrm{d}a)\sin^2\varphi - \mathrm{d}a.
\end{aligned} \tag{8.39}$$

Here, $\mathrm{d}X$, $\mathrm{d}Y$, and $\mathrm{d}Z$ are the geocentric translation parameters, M and N are the meridian and normal radii of curvature at the given latitude φ on the first ellipsoid, $\mathrm{d}a$ is the difference in the semimajor axes of the target and source ellipsoids, and $\mathrm{d}f$ is the difference in flattening of the two ellipsoids.

8.8.4 Projections and Plane Coordinates

GIS, cartography, and surveying applications need plane coordinates. There are numerous ways to project the ellipsoid onto the plane, but it is not possible to avoid distortions in distances in this process [32]. Either areas or angles can be selected as the target quantity that should not be distorted so as to obtain equal-area or conformal (equal angle) mapping in the plane. In geodesy and GIS, conformal maps are preferred, as the distance distortions are not dependent on the direction. They are computed from the solution of the Cauchy–Riemann differential equation (8.42). There are many possible choices of conformal maps. The modern form of the transverse Mercator projection, one of the most important conformal maps in GIS and geodesy, will be derived below. A comprehensive collection of formulas and parameters can be found in [33, 34]. A general feature of projections is that they are basically two-dimensional mappings between two surfaces, so heights above these surfaces remain unaffected.

Conformal mapping is facilitated by the one-to-one transformation of an isothermal net of parameter lines from one surface to the other. Here, *isothermal* means that both sets of parameter lines are orthogonal and of the same scale (isometric). The plane Cartesian coordinate set is isothermal; the net of meridians and parallels (i. e., the geographic longitude λ and the latitude φ) are not. However, an isothermal net on the sphere or the ellipsoid can be generated by using the Mercator function

$$q = \ln\left[\tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)\left(\frac{1 - e\sin\varphi}{1 + e\sin\varphi}\right)^{\frac{e}{2}}\right] \tag{8.40}$$

to convert the latitude φ to the isothermal latitude q. The inversion is done iteratively using $\varphi = 0$ as a starting value

$$\varphi_{i+1} = 2\arctan\left[\left(\frac{1 - e\sin\varphi_i}{1 + e\sin\varphi_i}\right)^{\frac{e}{2}} \cdot \exp(q)\right]. \tag{8.41}$$

Conformal mappings are now computed via the holomorphic function

$$x + \mathrm{i} \cdot y = f(q + \mathrm{i} \cdot \lambda), \tag{8.42}$$

where the plane coordinate system components x and y are the real and *imaginary parts* of the complex function (8.42), respectively.

8.8.5 Meridian Strip Projection (Transverse Mercator Projection)

This projection dates back to Gerhard Mercator in 1569, and was later derived by Gauss for ellipsoidal coordinates. The formulas used for practical calculations were developed by various geodesists in different countries, and this projection is known by several names, although the internationally accepted name is the *transverse Mercator projection* (TMP). Maps projected by the TMP have straight lines for all meridians and parallels. A particular meridian of the ellipsoid is adopted as the central meridian (CM) of the projection. Usually, the chosen central meridian in its true length, i.e. with a constant scale, has a longitude that is divisible by 3, i. e., 6°, 9°, 12°. A small region $\pm\Delta\lambda$ east and west of the CM is then mapped using (8.42) so that the projection of the CM becomes the x or north axis. The projection of the equator is the y- or east axis (Fig. 8.17). The origin of the plane coordinate system is thus the intersection of the CM with the equator of the reference ellipsoid at the chosen reference meridian. The distortion increases with the distance from the CM. Therefore, $\Delta\lambda$ is usually limited to 1.5° or 3°, leading to stripes of width 3° or 6°.

Practical formulas for the conversion that are based on series expansions have been published [35]. An analytical solution based on the complex variables of the holomorphic function (8.42) that is easily implemented was published in [36] and was elaborated upon in [37]. It is facilitated in a two-step procedure for forward and backward conversion from ellipsoidal to plane coordinates.

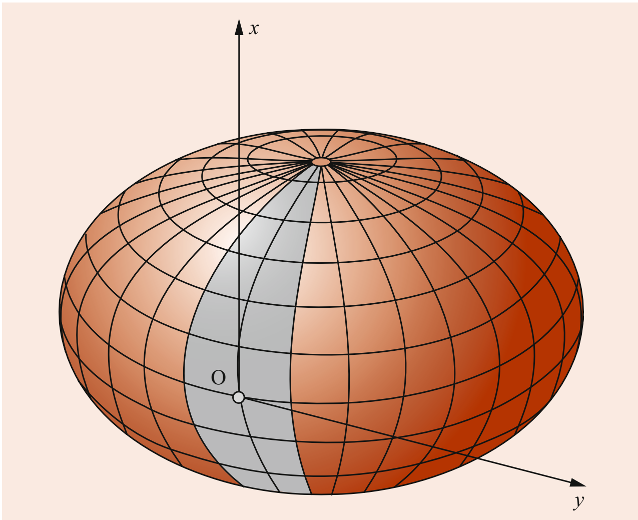


Fig. 8.17 Meridian strip projection scheme

Conversion Between (φ, λ) and (q, l)

$$\begin{aligned} q &= \operatorname{arctan} h(\sin\varphi) - e \operatorname{arctan} h(e \sin\varphi), \\ l &= \lambda, \\ \varphi_{i+1} &= \operatorname{arcsintan} h(q) + e \operatorname{arctan} h(e \sin\varphi_i), \\ \varphi_1 &= 0, \\ \lambda &= l. \end{aligned} \tag{8.43}$$

Conversion Between (q, l) and (x, y)

Here, the transformation between the two sets of isothermal coordinates is solved by the complex function $z = f(w)$, where the analytical function f is determined by presupposing an undistorted CM.

$$\begin{aligned} w &= q + \mathrm{i} \cdot l \\ z &= x + \mathrm{i} \cdot y \\ b_{i+1} &= \operatorname{arcsintan} h[w + e \operatorname{arctan} h(e \sin b_i)] \\ b_{\mathrm{i}} &= \left(\overline{\varphi} + \mathrm{i} \cdot \overline{\lambda}\right) \\ z &= M_0(1 + E) \cdot b - M_0 \frac{1}{2} \sin(2b) E_b, \text{ with} \\ E(e'^2) &= \sum_{n=1}^{\infty} d_n'' \\ E_\varphi(e'^2, \varphi) &= \sum_{n=1}^{\infty} d_n'' \sum_{k=0}^{n-1} k_n'' \\ d_{n+1} &= d_n \frac{(2n+1)(2n+3)}{(2n+2)(2n+2)} e^2; d_0 = 1 \\ k_{i+1} &= d_n \frac{(2i+2)}{(2i+3)} \sin^2\varphi; k_0 = 1. \end{aligned} \tag{8.44}$$

The plane coordinates (x, y) are, respectively, the real and imaginary parts of the complex variable z. In many countries, a constant (false easting (FE) and false northing (FN)) is added to the y coordinate in order to avoid a negative sign and to indicate the zone or chosen CM. Frequently, the resulting coordinates are then termed *right* or *east* for y and *high* or *north* for x. Of course, when performing the inverse transformation $x, y \rightarrow \varphi, \lambda$, this constant has to be eliminated first. In addition, the CM may not be projected isometrically but it must be multiplied by a scale factor, just as for UTM coordinates.

To achieve the inverse transformation, the above formulas are inverted, so q and l, respectively, are the real and imaginary parts of w.

Conversion Between (q, l) and (x, y)

$$\begin{aligned} b_{i+1} &= \frac{z}{M_0(1+E)} + \sin(2b_i) \frac{E b_{(i-1)}}{2(1+E)}; \; b_0 = 0, \\ w &= \operatorname{arctan} h(\sin b) - e \operatorname{arctan} h(e \sin b), \\ &\text{with} \quad M_0 = a(1 - e^2). \end{aligned} \tag{8.45}$$

Convergence is achieved quickly with these formulas; only three or four iterations at most are generally needed to obtain centimeter accuracy.

Distance and Area Distortions

Conformal mapping leaves angles unchanged, but distances and areas are increasingly distorted the farther they are from the CM. In principle, distances and areas are enlarged by the projection. Let s be the length of a geodetic line on the ellipsoid, s' the corresponding length in the plane, A the area on the ellipsoid, and A' the area in the plane. To get a sufficiently good approximation of the distortions, it suffices to use an osculating sphere of radius R_{m}, where R_{m} is the mean radius at the midlatitude of the region in question. Then, with y_1 and y_2 as the y coordinates of the endpoints of a line or the nearest and farthest points of an area with respect to the CM,

$$\begin{aligned} s' &= s + \frac{s}{6R_{\mathrm{m}}^2}\left(y_1^2 + y_1 y_2 + y_2^2\right), \\ A' &= A + \frac{A}{3R_{\mathrm{m}}^2}\left(y_1^2 + y_1 y_2 + y_2^2\right), \\ R_{\mathrm{m}}^2 &= \frac{c^2}{V_{\mathrm{m}}^4}. \end{aligned} \tag{8.46}$$

8.8.6 Universal Transverse Mercator System

To limit distortions, a grid system with several grid zones and common defining parameters is used. Coordinates throughout the system are repeated in each zone. To make coordinates unambiguous, the easting is prefixed by the relevant

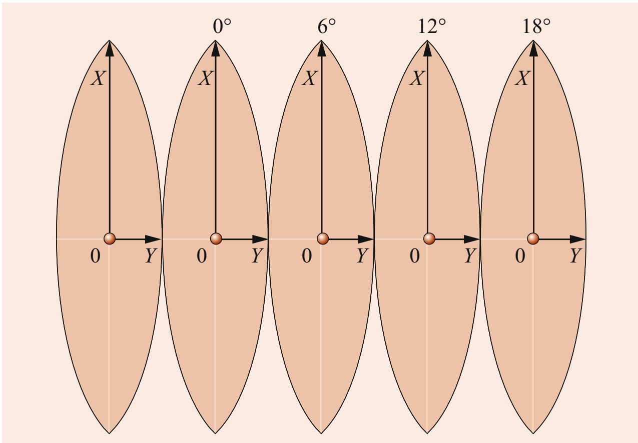


Fig. 8.18 Meridian strip projection scheme for UTM coordinates

zone number. This procedure was adopted, e. g., in Germany through the Gauss–Krüger system and later by the US military through the Universal Transverse Mercator (UTM) grid system (Fig. 8.18). The CM is a derived parameter computed from two other defining parameters: the initial longitude (the western limit of zone 1) (λ_I) and the zone width (W). Each of the remaining four transverse Mercator defining parameters—CM or latitude of natural origin, scale factor at natural origin, false easting, and false northing—has the same value in every zone.

The standard transverse Mercator formulas above are modified as follows. The zone number $Z = \text{INT}[(\lambda + \lambda_I + W)/W]$ with λ, λ_I, and W in degrees; λ_I is the initial longitude of the zoned grid system and W is the width of each zone of the zoned grid system. If $\lambda < 0$, $\lambda = (\lambda_I + 360)°$. Then

$$\lambda_0 = (ZW) - \left[\lambda_I + \left(\frac{W}{2}\right)\right]. \tag{8.47}$$

For the forward calculation, the easting and northing are

$$\text{E} = Z \times 106 + \text{FE} + k_0 y, \tag{8.48}$$

$$\text{N} = \text{FN} + k_0 x. \tag{8.49}$$

In the reverse calculation,

$$x = \frac{\text{N} - \text{FN}}{k_0}, \tag{8.50}$$

$$y = \frac{\text{E} - (\text{FE} + Z \times 106)}{k_0}. \tag{8.51}$$

In the UTM coordinate system introduced by the US military and now being used as standard in many countries, the Earth is divided into 60 zones, each with a longitudinal extension of 6° (Fig. 8.18). The zones are numbered from 1 to 60 beginning with the zone between 180° and 174° west of Greenwich and progressing eastward. The central meridians λ_0 are −177°, −171°, −165°, and so forth. The defining parameters are $\lambda_I = 177°$, $k_0 = 0.9996$, FE = 500 000 m, FN = 0 m north of equator or FN = 10 000 000.0 m from 80° southern latitude to equator, and $W = 6°$.

The use of the scale factor k_0 causes isometry to be lost for the CM but gained for two curves parallel to the CM at a distance of about ±180 km. Note that distortions within these two curves are negative, whereas distortions outside them are positive.

The UTM system is used only up to latitudes of 80° N and down to 80° S. This is because the convergence of the meridians causes rapidly varying zones when approaching the poles. For those regions, the polar stereographic projection (which is also conformal) is used. This is a projection onto a plane tangent to the ellipsoid at either of the poles. The origin of the plane system is the pole, and the images of the meridians are straight lines through the pole. Parallels of equal latitude are concentric circles.

Another often-used projection is the Lambert conformal conic projection, which is also used in international aeronautical charts.

8.8.7 Datum Transformation

The geodetic datum includes all information necessary to define a geodetic system. Considering the dimensions of the reference ellipsoid, the seven parameters for the transformation to the ITRF, and the height reference, ten parameters in total are required to define the geodetic datum unambiguously. These parameters are the three translations of the origin, three rotations around the Cartesian axes, a scale factor, and the two reference ellipsoid parameters a and f. In addition, for the vertical reference system of the heights, the potential of the geoid W_0 has to be specified. In the classical case of separate position and height coordinates, such as with two-dimensional triangulation networks, the datum has a slightly different definition. In that case, it comprises a reference surface consisting of the following parameters: the latitude and longitude of an initial point (origin), the orientation of the network, and the two parameters of a reference ellipsoid. A map projection needs information about a reference surface that is fixed by a geodetic datum in space (superior coordinate reference system). National mapping agencies define their own reference systems. Different countries may use the same parameters for the reference surface but with different positions and orientations. National mapping agencies may use different map projections that are based on the same reference system.

Plane coordinates for a particular ground location and its height will vary based on both the datum used to produce a particular map or chart and the design parameters of the map projection, like scale, area and anticipated distortion. Consequently a change of coordinates to another reference

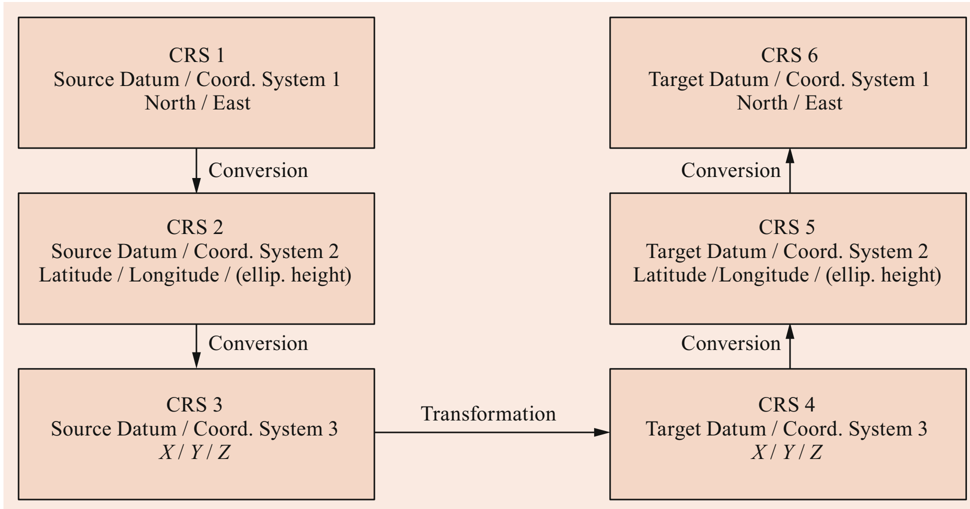


Fig. 8.19 Example of the combined coordinate conversion and transformation procedure (after [38])

system or a different map projection should always include both a geodetic datum transformation and the change of the map projection. Therefore, it is essential to include the datum parameters used to derive the coordinates when reporting positions. ITRS and the GRS80 (or equivalently, but less accurately, the WGS 84) now provide the single standard reference datum for geographic reference systems worldwide. They should be used in all new coordinate determinations and GIS applications. When unifying older data, respective transformations of the underlying reference systems and datums have to be performed based on the formulas described above. The recommended general procedure for transformation and conversion is now described.

The change of coordinates from one CRS to another often involves several steps. A scheme that involves a series of operations consisting of one or more transformations and one or more conversions is given in Fig. 8.19. If the necessary information is available, the switch from plane coordinates in the source system CRS1 to plane coordinates in the target system CRS6 should ideally be made using Cartesian ITRF coordinates to avoid distortions. For a rigorous transformation, the use of ellipsoidal coordinates of CRS2 and CRS5 is required. Databases of the parameters are compiled by various organizations [34]. However, for historical data, the ten parameters required are often not available, in which case it may be appropriate to use two- or three-dimensional similarity transformations based on adjustments performed over identical points. Missing height values, missing information on the type of or reference for the heights, and missing information on the ellipsoid and the datum of the triangulation are potential sources for major distortions and errors in the transformed coordinates, and sophisticated methods have to be applied.

Table 8.3 presents some examples of datum transformation parameters for illustration purposes. The origins of the historical systems prior to satellite geodesy all have large offsets and differences in the semimajor axis a and flattening f. Modern CRSs such as the WGS 84 are geocentric and the parameters depend on the precision of coordinate determination. For the most recent WGS 84 realization and the various ITRF realizations, the transformation parameters reflect improvements in reference frame determination over time.

Actual parameters should be taken from [8, 18].

8

8.9 Coordinate Determination

The geocoding of objects or features is done using geodetic techniques. There are two different approaches: space-based coordinate determination by GNSS observations and ter-

Table 8.3 Examples of historical and actual datum transformation parameters to ITRF2007 and the GRS80 ellipsoid

Datum	Ellipsoid	da (m)	df (10^{-3})	dX (m)	dY (m)	dZ (m)	M (10^{-9})	ω_x ($''$)	ω_y ($''$)	ω_z ($''$)
European 1950	International 1924	251	141.92702	−87	−96	−120	–	–	–	–
North American 1927	Clarke 1866	69.4	372.64639	−9	161	179	–	–	–	–
Pulkovo 1942	Krassovsky 1940	108	4.80795	28	−121	−77	–	–	–	–
ITRF 2005	GRS 80	–	–	0.03	0.04	−0.08	0.59	0.0	0.0	0.0
ITRF 97	GRS 80	–	–	0.02	0.03	−0.06	0.43	–	–	–
WGS 84 since 1994	WGS 84	–	0.00016	0.01	−0.01	0.02	7.7	0.003	0.000	−0.003
WGS 84 before 1994	WGS 84	–	0.00016	−0.060	0.52	0.22	0.01	−0.0183	0.0003	−0.007

Table 8.4 GNSS observables and typical accuracies for observations and coordinates

Observation type	Range accuracy (m)	Typical coordinate accuracy (m)
Code observations at one frequency	10–30	10
Code observations at one frequency supplemented by augmentation system information such as EGNOS in Europe or WAAS in the USA	5–10	5
Dual-frequency code observations	2–5	2–3
Carrier phase observations at two frequencies	0.003	0.01–0.03

restrial geodetic methods such as tacheometry. Terrestrial methods primarily result in coordinates in a local topocentric system that are relative to an existing geodetic reference point. GNSS coordinate results are in absolute global coordinates, e. g., in the WGS 84, the ITRF, or a national datum. The principles of these two different approaches are explained in the next two subsections.

8.9.1 GNSS Coordinate Determination

GPS can be used as a proxy for all future GNSSs, such as the Russian GLONASS, the European Galileo system, and the Chinese BeiDou system. Although originally built for military purposes, civilian use is guaranteed by the operating governments. The space-based component of each of the systems, a set of 24–30 satellites, orbits the Earth at an average height of 20 000 km in a time of about 12 h. These orbits are very precisely known from tracking network observations. Each of the satellites transmits information on its position at two or more microwave frequencies; these are 1575.42 MHz (L1) and 1227.6 MHz (L2) for GPS. The carrier frequency is modulated by codes so that each satellite can be identified and the distance between the satellite and the user receiver–antenna can be calculated. This distance is called the *pseudorange*, as it is derived from the travel time computed from the precise timing signals included in the data message. The user's position is determined from the known positions of at least four satellites and their respective ranges. As the orbits are given in the ITRF, the user coordinates are also in this global geocentric reference frame. In practice, most modern receivers offer a coordinate conversion, e. g., to ellipsoidal coordinates and height (ellipsoidal or above sea level if the EGM model is considered). Alternatively, projected plane coordinates in a specified datum are often implemented.

There are different levels of accuracy that can be achieved by GNSS observations depending on the observables used. Four main levels may be distinguished, as given in Table 8.4.

The receivers needed for the different levels differ markedly in price. The dual-frequency and phase observables options usually associated with geodetic receivers are the most expensive. The use of phase observables is the most advanced technique and requires sophisticated software for evaluation due to the ambiguity inherent in the phase observations. The distance is composed of an integer number of full-wavelength cycles of the carrier wave plus an actually observed phase value. The integer number of cycles cannot be determined directly by the receiver, and therefore longer observing times or advanced filtering techniques are needed to resolve these ambiguities. Only after applying these procedures can an accuracy of a few centimeters be obtained. When the phase observables are used in combination with the code observations, *precise point positioning* (PPP) can be achieved. This is becoming a new standard due to the new generation of multi-GNSS receivers with multifrequency code and phase observations. For further information on the use of GNSSs, refer to [39].

The resultant positional accuracy is essentially determined by disturbances to the satellite clock, in the atmosphere, and to the geometric constellation of the satellites used. The expected accuracy from GNSS observations can be estimated in advance using the dilution of precision value (DOP). There are several versions of this; for point positioning, the geometric dilution of precision (GDOP) and the position dilution of precision (PDOP) are the most relevant. The GDOP describes the quality of the solution and is proportional to the volume generated by the polyhedron formed by the satellites and the user's antenna. Both the PDOP and the GDOP can be computed in advance without actual observations from approximate coordinates and the predicted satellite ephemeris in order to, e. g., plan the optimal time for a measurement. The PDOP is computed from the trace of the cofactor matrix of the adjustment of the position. Having conducted observations, the accuracy is obtained from the covariance matrix of the final adjustment results of the position computation. GDOP values less than or equal to 4 are well suited for positioning, whereas values larger than 8 should be avoided. For actual surveys, the visibility of the sky should be such that there is a homogeneous distribution of more than four satellites at elevations larger than 10°. Special attention is required in scenarios where satellites are blocked by buildings, urban canyons or in forest areas. Here, reflected signals or degraded satellite geometry may cause larger errors.

Relative Positioning

Differential observations are often used to mitigate the influence of the high spatial correlations of the main GNSS error sources. This mode of operation is called differential GNSS (DGNSS). A reference receiver is placed at a base station with a known position to log the GNSS data. One or more roving receivers are used for new points. The difference between the known and actual determined positions at the base station is used to compute corrections to the observables. These corrections—or the raw observation data from the base receiver—can be transmitted by a radio link or via the Internet to the rover and evaluated in real time or in postprocessing. There are three basic types of differential corrections:

1. Corrections to the coordinates determined by the rover. These are associated with the National Marine Electronics Association (NMEA) format standard NMEA-183.
2. Corrections to the pseudoranges observed at the rover. These are associated with Radio Technical Commission for Maritime Services (RTCM) or Radio Technical Commission for Aeronautics (RTCA) standards such as RTCM-10403.1.
3. Raw phase data or corrections to phase data. These are associated with RTCM version 2.3 or version 3 standards (REF RTCM).

The improvement in accuracy achieved by using differential GNSS depends on the distance between the base and the rover. In many countries, commercial services that offer various corrections are available. They are based on dense networks of reference stations such as the German Satellitenpositionierungsdienst der deutschen Landesvermessung (SAPOS) or the National Geodetic Survey (NGS) Continuously Operating Reference Stations (CORS) in the US [40]. The accuracy is in the range of 3–6, 0.6–2, or, at best, 0.01–0.03 m for differential phase data usage.

GNSSs Today

The outlook for GNSS positioning is bright. The *system of systems* with four GNSSs and more than 120 satellites has dramatically increased the number of visible satellites from about 5–12 for a single GNSS to an average of about 30 satellites each day. This significantly improves the positional dilution of precision (PDOP), the measure of the geometric quality of the positioning solution. Intergovernmental agreements guarantee the interoperability and compatibility of the GNSSs. Compatibility refers to the ability of global and regional navigation satellite systems and augmentations to be used separately or together without causing unacceptable interference and/or other harm to an individual system and/or service [41]. Interoperability refers to the ability of global and regional navigation satellite systems and augmentations and the services they provide to be used together to provide better capabilities at the user level than would be achieved by relying solely on the open signals of one system. In practice, this means that modern receivers that use at least the basic frequency of 1575.42 MHz are able to process all systems individually or jointly; the latter case will provide quicker, more accurate, and more robust positioning solutions than any single system. This is of particular importance in kinematic applications and locations with reduced visibility such as urban canyons or in forests. A detailed study of the multi-constellation GNSS positioning performance can be found in [42]. Moreover, the new generation of GNSSs have three carrier frequencies, i. e., a third frequency at about 1189 MHz that is common to all GNSSs. New signal types with advanced modulations are being added to all carriers, meaning that dual frequency code observations will no longer be restricted to costly high-end geodetic receivers. Receivers capable of such observations are about to enter the mass market. New signals have a better signal-to-noise ratio and a smaller code multipath, facilitating less noisy and more robust positioning solutions and potentially opening the door to some applications with meter-level accuracy indoors.

Multi-GNSS and multifrequency availability has also boosted the use of PPP. There are a variety of commercial providers of PPP products that allow standalone kinematic positioning down to the centimeter level globally. Products such as precise orbits, clock corrections, and atmospheric state information are transmitted via the Internet or by dedicated communication satellites. Open services are coordinated by the IGS Multi-GNSS pilot project (MGEX) [43], which allows the use of correction data based on dedicated servers [44]. Fields of application are typically kinematic applications such as precision farming, positioning for simultaneous location and mapping, and static applications as an alternative to classical differential or global positioning. Present performance parameters for GPS—still the best single system due to its full constellation—include an accuracy of around 10 cm after 30 min of convergence time. After 2 h, the 5 cm level can be obtained, and the static daily session accuracy is close to that of a classical solution, at the millimeter level. Multi-GNSS PPP yields an accuracy after only 5 min of convergence time of better than 10 cm and significantly improved robustness [42]. Moreover, combining the RTK and PPP techniques allows for the use of low-cost single-frequency receivers or even smartphones for centimeter-level positioning in, e. g., automotive applications [45, 46].

8.9.2 Terrestrial and Local Coordinate Determination

Terrestrial measurements are made in the local topocentric coordinate system. Terrestrial techniques are used in areas with low GNSS observability and short distances (in the kilometer range). Instruments are oriented along the local vertical, so they are related to the actual gravity field at the observation site. The slant distances and horizontal and vertical angles measured by tacheometers allow polar coor-

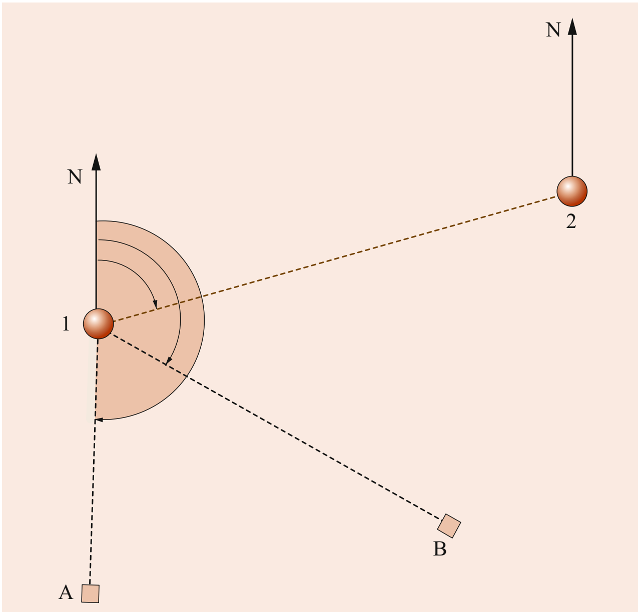


Fig. 8.20 Polar coordinate determination; *rectangles* are known points, *circles* are new points

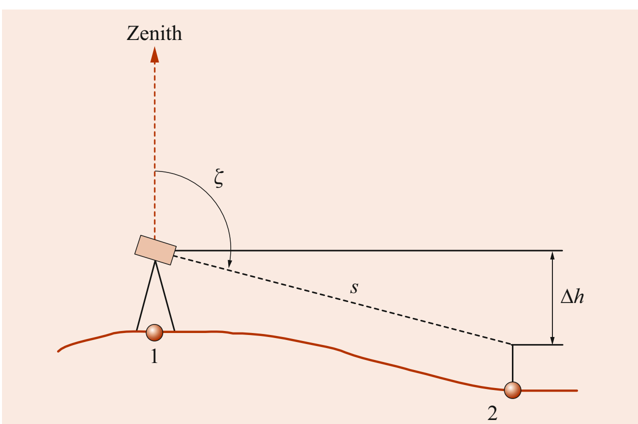


Fig. 8.21 Measurement of a height difference by a tachometer: distance and zenith angle

dinates to be determined and new points to be attached to known reference points. The combination of a GNSS with a tacheometer is often practical. This requires at least one known point to set up the instrument and another for the direction reference to the grid coordinates (Fig. 8.20). The distance and the vertical angle (zenith angle) are used to determine the height (Fig. 8.21). Local observations have to be corrected for the local distortion and are then used to compute grid coordinates, e. g., UTM northing and easting. Alternatively, the local 3-D vector can be transformed into geocentric ITRF Cartesian coordinates using (8.31).

Modern tacheometers are available with several levels of accuracy; typical instruments measure distances to the level of $0.003 \pm 2 \times (10^{-6} \times \text{distance})$ in meters. Given that the uncertainty in angular measurements is less than 4 arcsec, coordinates can be determined with uncertainties of 0.01–0.03 m at distances of a few km. Tacheometer packages facilitate full integration with GNSSs using a software suite, object coding for a GIS using an electronic field book, and a database connection. Mobile GIS applications allow the formation of objects from line and area data and the possibility of object classification with descriptive data in the field.

References

1. Beutler, G., Rothacher, M., Schaer, S., Springer, T.A., Kouba, J., Neilan, R.E.: The International GPS Service (IGS): An interdisciplinary service in support of Earth sciences. Adv. Space. Res. **23**, 631–653 (1999)
2. Dow, J.M., Neilan, R.E., Rizos, C.: The International GNSS Service in a changing landscape of Global Navigation Satellite Systems. J. Geod. **83**, 191–198 (2009)
3. Teunissen, P.J.G., Montenbruck, O. (eds.): Springer Handbook of Global Navigation Satellite Systems. Springer, Cham (2017)
4. Li, X., Ge, M., Dai, X., Ren, X., Fritsche, M., Wickert, J., Schuh, H.: Accuracy and reliability of multi-GNSS real-time precise positioning: GPS, GLONASS, BeiDou, and Galileo. J. Geod. **89**, 607–635 (2015)
5. Rummel, R.: Integrated Global Geodetic Observing System (IGGOS). J. Geodyn. **40**, 357–362 (2005)
6. Seitz, F., Schuh, H.: Earth rotation. In: Xu, G. (ed.) Sciences of Geodesy – I. Advances and Future Directions. Springer, Berlin, Heidelberg (2010)
7. IERS: International Earth Rotation and Reference Systems Service. Agency for Cartography and Geodesy, Frankfurt (2010). http://www.iers.org/ (last accessed 30/01/2018)
8. ITRF: International Terrestrial Reference Frame. Institute Geographique National, Paris (2016). http://itrf.ign.fr/
9. Fey, A.L., Gordon, D., Jacobs, C.S., Ma, C., Gaume, R.A., Arias, E.F., Bianco, G., Boboltz, D.A., Böckmann, S., Bolotin, S.: The second realization of the international celestial reference frame by very long baseline interferometry. Astron. J. **150**(2), 58 (2015). https://doi.org/10.1088/0004-6256/150/2/58
10. Eurogeographics: http://www.eurogeographics.org (Champs-sur-Marne 2010).
11. IERS: Conventional Terrestrial Reference System and Frame, IERS Conventions. IERS, Sevres (2007). Chap. 2
12. Kovalevsky, J.: The Hipparcos Catalogue as a Realisation of the Extragalactic Reference Frame. Astron. Astrophys. **323**, 620–633 (1997)
13. Altamimi, Z., Rebischung, P., Métivier, L., Collilieux, X.: ITRF2014: A new release of the International Terrestrial Reference Frame modeling nonlinear station motions. J. Geophys. Res. Solid Earth **121**, 6109–6131 (2016)
14. DeMets, C., Gordon, R.G., Argus, D.F., Stein, S.: Effect of recent revisions to the geomagnetic reversal time scale on estimates of current plate motions. Geophys. Res. Lett. **21**(20), 2191–2194 (1994)
15. Ihde, J., Luthardt, J., Boucer, C., Dunkley, P., Farrell, B., Gubler, E., Torres, J. (eds.): European Spatial Reference Systems, Frames for Geoinformation Systems. CRS-EU Operator/Bundesamt für Kartographie und Geodäsie, Leipzig (1999)
16. Altamimi, Z.: EUREF Technical Note 1: Relationship and Transformation between the International and the European Terrestrial Reference Systems. Institut National de l'Information Géographique et Forestière (IGN), Paris (2017). http://etrs89.ensg.ign.fr/pub/EUREF-TN-1.pdf

17. Pearse, M.B.: Realisation of the New Zealand Geodetic Datum 2000. Office of the Surveyor-General Technical Report No. 5. Land Information New Zealand, Wellington (2000). http://www.linz.govt.nz/geodetic/standards-publications/technical-reports/index.aspx (last accessed 03/10/2010)
18. NIMA: Department of Defense World Geodetic System 1984: Its definition and relationships with local geodetic systems. Tech. Rep. TR8350.2. National Imagery and Mapping Agency NIMA, Bethesda (1997)
19. Moritz, H.: Geodetic Reference System of 1980 (GRS80) Bulletin Géodésique. In: Andersen, O.B. (ed.) The Geodesists Handbook 1969. International Union of Geodesy and Geophysics, Karlsruhe (1988). http://www.gfy.ku.dk/~iag/handbook/geodeti.htm (last accessed 03/10/2010).
20. Hofmann-Wellenhof, B., Moritz, H.: Physical Geodesy, 2nd edn. Springer, Berlin, Heidelberg (2006)
21. Kenyeres, A.: GPS/Leveling. In: Grafarend, E. (ed.) Encyclopedia of Geodesy, pp. 1–3. Springer, Cham (2016)
22. Bretterbauer, K.: A Primer of Geodesy. TU Vienna, Vienna (2003)
23. Sacher, M., Liebsch, G., Ihde, J., Luthardt, J.: Steps on the way to UELN05 and enhancements of the web-based Geodetic Information and Service System (CRS-EU). Mitt. Bundesamt. Kartogr. Geod. **38**, 158–165 (2006)
24. Ihde, J., Sánchez, L., Barzaghi, R., Drewes, H., Foerste, C., Gruber, T., Liebsch, G., Marti, U., Pail, R., Sideris, M.: Definition and proposed realization of the International Height Reference System (IHRS). Surv. Geophys. **38**, 549–570 (2017)
25. Lemoine, F.G., Kenyon, S.C., Factor, J.K., Trimmer, R.G., Pavlis, N.K., Chinn, D.S., Cox, C.M., Klosko, S.M., Luthcke, S.B., Torrence, M.H., Wang, Y.M., Williamson, R.G., Pavlis, E.C., Rapp, R.H., Olson, T.R.: The Development of the Joint NASA GSFC and National Imagery and Mapping Agency (NIMA) Geopotential Model EGM96. NASA Goddard Space Flight Centre, Greenbelt (1998). NASA/TP-1998-206861
26. NGA: Earth Gravitational Model EGM2008. National Geospatial Agency, Bethesda (2008). http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008/index.html (last accessed 03/10/2010)
27. Torge, W.: Geodesy. deGruyter, Berlin (2003)
28. NGA-GIS: Earth Gravitational Model EGM2008-GIS. National Geospatial Agency, Bethesda (2008). http://earth-info.nima.mil/GandG/wgs84/gravitymod/egm2008/egm08_gis.html (last accessed 03/10/2010)
29. Barthelmes, F., Köhler, W.: International centre for global earth models. J. Geod. **90**(10), 907–1205 (2016)
30. Guinot, B., Arias, E.F.: Atomic time-keeping from 1955 to the present. Metrologia **42**, 20–29 (2005)
31. Moudrak, A., Eissfeller, B., Hein, G., Klein, H.: Future time. Opportunities for using optical clocks in GNSS systems. Inside GNSS **9/10**, 45–50 (2008). http://www.insidegnss.com/node/768 (last accessed 03/10/2010).
32. Hooijberg, M.: Geometrical Geodesy. Springer, Berlin, Heidelberg, p. 439 (2008)
33. Snyder, J.P.: Map Projections – A Working Manual. Professional Paper, vol. 1395. US Geological Survey, Reston (1987). http://pubs.er.usgs.gov/publication/pp1395 (last accessed 03/10/2010)
34. OGP: Coordinate conversions and transformations including formulas. Geomatics Guidance Note, vol. 7. OGP Geomatics Committee, London (2011). Pt. 2, OGP Publication 373-7-2
35. Hooijberg, M.: Practical Geodesy. Springer, Berlin, Heidelberg (1997)
36. Klotz, J.: Eine analytische Lösung der Gauß-Krüger-Abbildung. Z. Vermessungswes. **118**(3), 106–116 (1993)
37. Stuifbergen, N.: Wide Zone Transverse Mercator Projection. Can. Tech. Rep. Hydrogr. Ocean Sci., vol. 262. Canadian Hydrographic Service, Dartmouth (2009)
38. CRS: Coordinate Reference Systems in Europe. CRS-EU Operator/Bundesamt für Kartographie und Geodäsie, Leipzig (2010). http://www.crs-geo.eu (last accessed 03/10/2010)
39. Hofmann-Wellenhof, B., Lichtenegger, H., Wasle, E.: GNSS Global Navigation Satellite Systems. Springer, Berlin, Heidelberg (2008)
40. CORS: Continuously Operating Reference Stations. National Geodetic Survey, Bethesda (2010). http://www.ngs.noaa.gov/CORS/cors-data.html (last accessed 03/10/2010)
41. IGC: Current and Planned Global and Regional Navigation Satellite Systems and Satellite-based Augmentation Systems. ICG Forum, UN Office for Outer Space Affairs, Vienna (2010). http://www.oosa.unvienna.org/pdf/publications/icg_ebook.pdf
42. Li, X., Zhang, X., Ren, X., Fritsche, M., Wickert, J., Schuh, H.: Precise positioning with current multi-constellation Global Navigation Satellite Systems: GPS, GLONASS, Galileo and BeiDou. Sci. Rep. **5**, 8328 (2015)
43. Montenbruck, O., Steigenberger, P., Prange, L., Deng, Z., Zhao, Q., Perosanz, F., Romero, I., Noll, C., Stürze, A., Weber, G., Schmid, R., MacLeod, K., Schaer, S.: The Multi-GNSS Experiment (MGEX) of the International GNSS Service (IGS) – Achievements, prospects and challenges. Adv. Space. Res. **59**(7), 1671–1697 (2017)
44. BKG: Real time GNSS. Bundesamt für Kartographie und Geodäsie (BKG) Data Center (2018). https://igs.bkg.bund.de/ntrip
45. Mongredien, C., Hide, C., Fairhurst, P., Ammann, D.: Centimeter positioning for UAVs and mass-market applications. GPS World **25**, 24–33 (2016)
46. Murrian, M.J., Gonzalez, C.W., Humphreys, T.E., Pesyna, K.M., Shepard, D.P., Kerns, A.J.: Low-cost precise positioning for automated vehicles. GPS World **27**(9), 32–39 (2016)

Matthias Becker Professor Becker received his PhD in Geodesy in 1984. Since 2000 he has chaired the Applied Geodesy Department at the Universität der Bundeswehr München. Since December 2005 he has been Director of the Institute for Physical Geodesy at the Technische Universität Darmstadt. His present research activities are precise positioning by GNSS and Galileo, mass transport and mass distribution in the Earth system, and geodynamics.

8

9 Data Acquisition in Geographic Information Systems

Jan Skaloud, Michael Cramer, and Norbert Haala

Contents

J. Skaloud (✉)
Swiss Federal Institute of Technology Lausanne (EPFL)
Lausanne, Switzerland
e-mail: jan.skaloud@epfl.ch

M. Cramer
Institute for Photogrammetry (ifp), Stuttgart University
Stuttgart, Germany
e-mail: michael.cramer@ifp.uni-stuttgart.de

N. Haala
Institute for Photogrammetry (ifp), University of Stuttgart
Stuttgart, Germany
e-mail: norbert.haala@ifp.uni-stuttgart.de

Abstract

Geographic information systems (GIS) receive data from many sources that are different in technology, geographic coverage, date of capture, and accuracy – to mention few categories. The vast majority of the today's topographical and GIS-data are captured from mobile and possibly autonomous platforms that operate from the air, on the ground (also indoors) or on the water and that are equipped with optical sensors. Although the palette of optical sensors is rather large the most useful for mapping purposes falls into two categories. First are the passive sensors such as digital cameras in frame or line configuration. The main technological concepts of these sensors are introduced in Optical Sensors together with Lidar that serves the acquisition of detailed terrain structure in natural areas. The optical acquisition is supported by trajectory determination through the combined use of integrated navigation technology, which main concepts are outlined in Navigation Sensors. The geometrical principals of 3-D restitution of the scene are described first in Photogrammetry for the case of frame imagery only, later in Sensor Fusion for active sensors and integrated approaches. An overview of Mapping Products concludes this chapter.

Keywords

optical sensors · navigation sensors · photogrammetry · sensor fusion · mapping · adjustment · Lidar · INS/GNSS

9.1 Optical Sensors

The acquisition of surface texture is mainly obtained by passive optical sensors that originate from the principles of photography. Photography is a passive method, i.e., the solar energy reflected from the object (or the emitted thermal energy) is recorded by photosensitive materi-

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_9

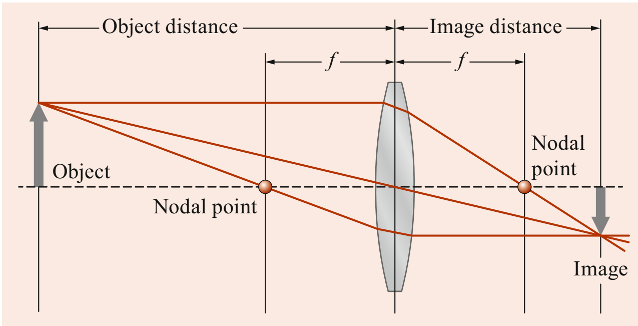


Fig. 9.1 Basic imaging principe when using a lens or lens system

als or elements. Modern electronic light-sensitive elements are charge-coupled (CCDs) or complementary metal-oxide-semiconductor (CMOS) devices. A CCD array consists of coupled detectors that allow charge to be moved across the array into capacitor bins for further processing. A CMOS detector works independently of neighboring detectors (pixels), as each one has an attached transistor that controls the analog-to-digital conversion and subsequent readout. A CMOS sensor is less expensive to manufacture and, principally, has a faster readout.

Photography in its simplest case is based on the pinhole camera model [1]. The geometric theory of optical systems assumes straight light rays that have been reflected from an object illuminated by any light source. These rays enter the camera through the pinhole, forming an inverted image on the plane opposite to the pinhole. This is where the photo sensitive material is placed. In the pinhole camera model, described in detail in Sect. 9.3.3, each image point is generated by one single light ray passing the pinhole. The resulting 2-D image is an ideal projection of the 3-D objects, since the simple pinhole model neglects, for example, any distortions and blurring effects due to defocus. It is assumed that pinhole, object point, and image point define one straight line (central perspective). The pinhole has the drawback that, ideally, only one single ray from the bundle of light rays originated from the object point source forms the corresponding image point. Thus, lenses, single or more complex lens systems, are used to enlarge the size of the camera opening but still retaining a focused image. The optical axis of such a lens system is defined as the line between incident and emergent nodal point, which are defined in a way that the chief or central rays pass the lens (system) without deviation, forming the same angles to the optical axis in both nodal points [2], i.e., the emerging ray is parallel to the original incident ray (Fig. 9.1).

The nodal points define the object–space incident and image-space emergent perspective centers. Light rays entering the lens parallel to the optical axis come into focus at the focal point. The plane perpendicular to the optical axis, including the focal point, defines the plane of infinite focus or simply the focal plane [1]. Any parallel rays entering the system come into focus in this plane. The focal length of the system is the distance between the emergent nodal point and the focal plane. The principal point is defined where the optical axis hits the focal plane. The principal point and focal length are the elementary geometrical parameters defining the geometry of a camera. This is called the interior (inner) orientation of the camera. If this interior orientation of the camera and its corresponding image is known, a bundle of image rays can be reconstructed from observed image coordinate measurements. The connection between the bundle of rays towards their correspondences in the image space is expressed by collinearity equation, which is presented in Sect. 9.3. Reconstruction of 3-D coordinates of objects derived exclusively from 2-D image observations is detailed in Sect. 9.3, while the benefit of integrating navigation sensors into this process is described in Sect. 9.4.

9.1.1 Frame Cameras

Until recently, the acquisition of texture was done exclusively using films. These have now been almost completely replaced by electronic sensors, at least for consumer-grade photography. Consequently, and similarly to consumer grade photography, analog film-based systems are rapidly being phased out in operational photogrammetric environments, and digital sensors have, to a large extent, replaced their analog predecessors. According to their geometry airborne digital cameras fall into the two large categories of frame and line cameras, where the latter are also referred to as push-broom sensors. The concepts of both of these are described in this and the following Sect. 9.1.2.

Single Head

Airborne or satellite platforms employ large image formats to guarantee an efficient data acquisition, as the available image size directly influences the cost of covering a certain area with imagery. Indeed, a smaller image format requires more images to record a given scene with the same spatial resolution. Especially in airborne imaging, this negatively influences the efficiency of image data recording and processing. Therefore, traditional analog mapping cameras have been designed with large formats of about $23 \times 23\,cm^2$. For these, focal lengths of 30, 15, or 8 cm are utilized, depending on the field of view (FOV) needed, which is 60, 95, and 125 degrees, respectively.

Multiple-Head Cameras

The most intuitive way to design a digital mapping camera would be to replace the former analog film by a 2-D electronic sensor element or sensor matrix. Indeed, such an approach has been pursued in consumer-grade photography.

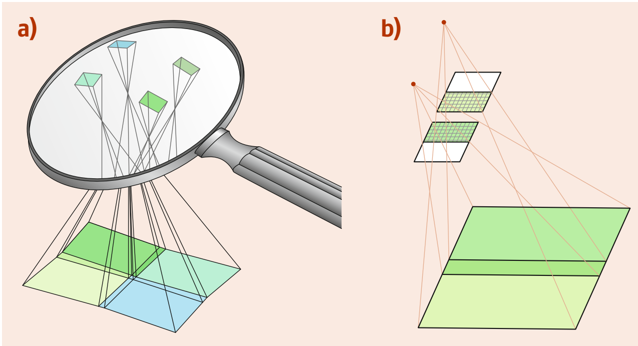


Fig. 9.2 Camera concepts based on CCD frame arrangements, synchronous recording: (**a**) "butterfly arrangement" with 4 tilted heads, each head with one CCD located in center position, (**b**) "shifted CCD arrangement", here with 2 nadir-looking heads, each with one CCD located in shifted position. Courtesy of ifp-Stuttgart

Unfortunately, the size of a CCD frame is physically limited by the supporting electronics. Therefore, was some years until a special multiple-head concept was developed based on a cluster of CCD sensors with a format comparable to the former 35 mm format (24 mm × 36 mm negative). Such a design employs several individual camera heads, each one equipped with one or more CCD frame sensors that are all firmly attached to one airborne platform. Due to the special geometrical arrangements, individual CCDs of a smaller format connected to separate camera heads generate multiple smaller-format images with certain overlaps. This allows for the generation of one synthetic large-format image afterwards, which is obtained by resampling the individual smaller-format single images to a virtual large format on one focal plane. In other words, the virtual large-format image can be used in later production in the same way as any other frame images. The only difference is its derivation from a virtual camera instead of a physically existing camera.

Often, multihead frame cameras are designed in a way that the individual heads (generally two or four) are arranged with slightly oblique viewing directions. Such inclined installation of camera heads results in four overlapping images, the so-called butterfly pattern, which is necessary to form a virtual image of a large format (Fig. 9.2). These overlaps are necessary for the later transfer of corresponding points (also called tie points) or tie features, which enables the merging of several smaller images into one virtual image of a larger format.

Different to the concept of tilted camera heads that generate overlapping images, some installations rely on nadir-looking camera heads only. One approach is to slightly shift the CCD frames against each other in the neighboring camera cones. There the CCDs are not placed in the center of each focal plane but slightly decentered, shifted to the direction of the opposite edges of the individual focal planes. As the images are taken at the same time, part of the covered scene overlaps and, therefore, the images can be merged together (Fig. 9.2).

Syntopic Frame

In the concepts of multihead frame cameras presented so far, the images originating from the individual cones are taken at the same time, i.e., the image exposure of individual camera heads is synchronized. A different concept relying on multiple nadir-looking cameras takes images at different times, however, over the same place. Such an approach is called syntopic image recording, and it is based on the idea that multiple camera cones are arranged in a line, which coincides with the main flight direction (Fig. 9.3). If the different camera heads take their images one after the other, and this time shift corresponds exactly to the velocity of the camera movement, the images will be taken at the same position. As a result, the camera stations for all images are the same. Different to the previously described system layout, the camera cones contain between one to four CCD frames, which are installed in different arrangements in their focal planes. Depending on the individual arrangement of CCD frames within the different camera heads, overlapping images are generated in the object space, which can again be merged together afterwards. The concept of image formation from syntopic imaging is illustrated in Fig. 9.3. As can be seen, up to four CCDs are placed within one single focal plane in a special pattern. The cone containing the four CCD sensors

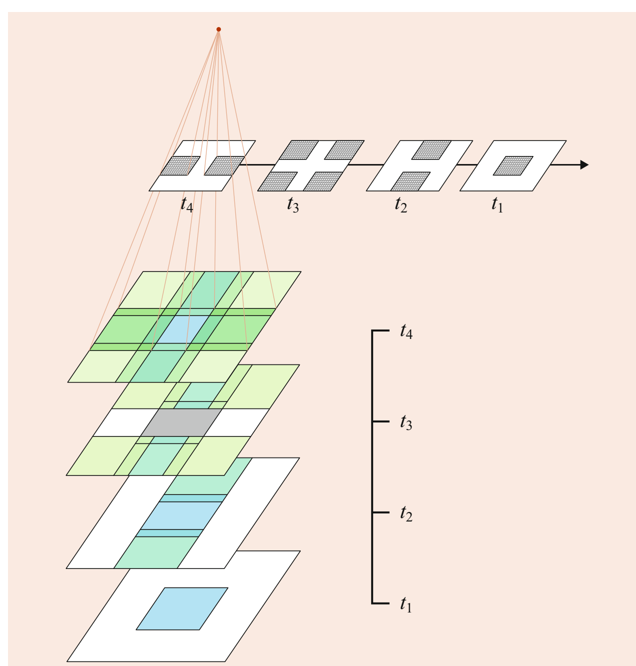


Fig. 9.3 The syntopic imaging concepts. Courtesy of ifp-Stuttgart

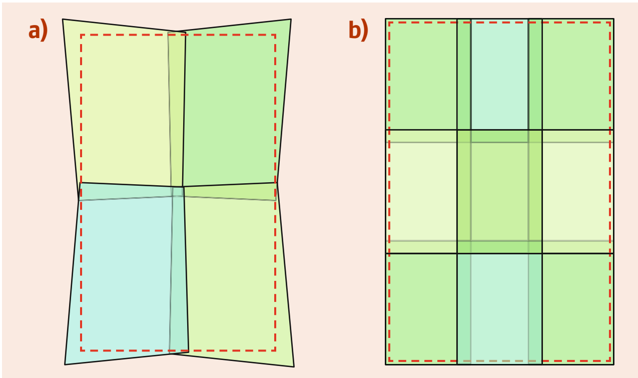


Fig. 9.4 Multihead concept to generate large-format frames: synchronous imaging using 4 tilted camera heads (**a**) and syntopic imaging using 4 nadir-pointing camera heads (**b**). Courtesy of ifp-Stuttgart

in its corners defines the virtual frame obtained after image stitching. It is named the primary or master cone. The remaining cones are used to fill the gaps in between.

Comparison of Concepts

Figure 9.4 compares the two ground patterns of the two main multihead concepts: four camera heads are used for both installations, but in the first case, the images are taken at the same time (synchronously) with tilted camera heads. Fig. 9.4a shows the particular footprint of such a setup in the object space, where the different colors indicate each of the four camera heads. Due to their off-nadir viewing the four images have individual perspective displacements. This tilt influences the imaging of the same objects in two camera heads, which is especially of concern in the overlapping parts. The effect is dependent on the height differences in object space but should be negligible in most application scenarios [3].

Syntopic image recording delivers a different pattern (Fig. 9.4b). Again, the color shades indicate the arrangement of CCD frames in the four participating camera cones. All cameras record at the same place (due to the small time interval between the different recordings) in the nadir-looking direction. Therefore, they should have the same perspective displacements, as long as the difference in the perspective center coordinates or off-nadir variations are negligible. Again, the overlapping regions between them are used for the formation of large-format imagery.

The dashed frame in Fig. 9.4 indicates those parts of the images that are used to form the virtual image of a large format. As can be seen, smaller parts at the corners of the butterfly pattern are lost. This is because the format of the virtual image is chosen to be rectangular. In the case of the syntopic imaging, the virtual image may use the full part of the individual frames. Practically, a small margin is also cut off in this approach.

Virtual Frame

The multihead concepts allow for the generation of virtual images of a larger format. For this purpose, several individual images are resampled to a previously defined virtual focal plane. This is based on the individual interior orientation of each camera head and their orientations relatively to each other. The process is called intercone orientation or *image stitching*. The knowledge of interior orientation (see Sect. 9.4 for definition) of every contributing camera head is necessary to exactly reconstruct the 3-D image rays originated from each camera-head pixel. Moreover, the relative orientation between the different camera heads (represented by six independent parameters) is required to determine their relation to the virtual focal plane. All together this defines the correct position where the image ray intersects the virtual large-format plane. The interior orientation of the camera heads is assumed to be known, whereas the orientation between the different camera modules is derived from conjugate points measured from the overlaps between the different images [3, 4]. Even in multihead cameras, where the different cones are mounted on one platform and images are taken in synchronized mode, the existence of such overlapping regions is necessary to control the stability of the orientation between the individual camera cones.

Color Generation

The generation of multispectral images from frame-based sensors can use several concepts. Many digital frame cameras, especially those designed for the consumer market, use the so-called Bayer pattern approach, where typically red, green, and blue (RGB) filters are arranged over every pixel on the CCD sensor in a special pattern. Thus, each pixel become sensitive only to one of the three base colors. The color is then derived through interpolation from neighboring pixels that contains the RGB components.

An alternative concept is to employ separate camera heads for each of the requested multispectral channels. Appropriate filters let each CCD array only capture the corresponding color information. Red, green, blue, and additional near-infrared spectral bands are the most common. Full RGB is derived through a so-called registration of color bands. The different images are overlaid to generate full color after adding three selected colors. In order to guarantee congruent features in the different color images a geometrical (2-D) transformation of images based on corresponding matched points between the different channels is necessary.

Figure 9.5 illustrates both concepts. When using the Bayer pattern approach, one CCD frame is sufficient to capture the full color information, but due to the pixel-wise color filters, each pixel only contains the color information of the corresponding channel. Additionally, 50% of all pixels are sensitive to the green spectral band while only 25% are sen-

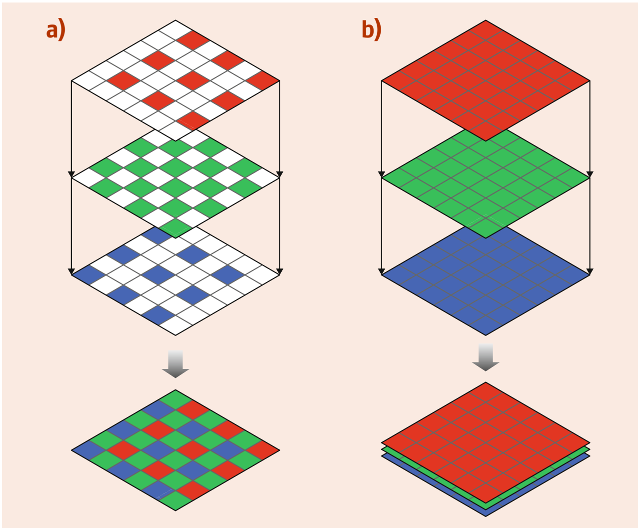


Fig. 9.5 The two concepts to obtain colored RGB digital images from frame sensors: (**a**) single CCD using the Bayer pattern. (**b**) individual CCDs, one for each color channel

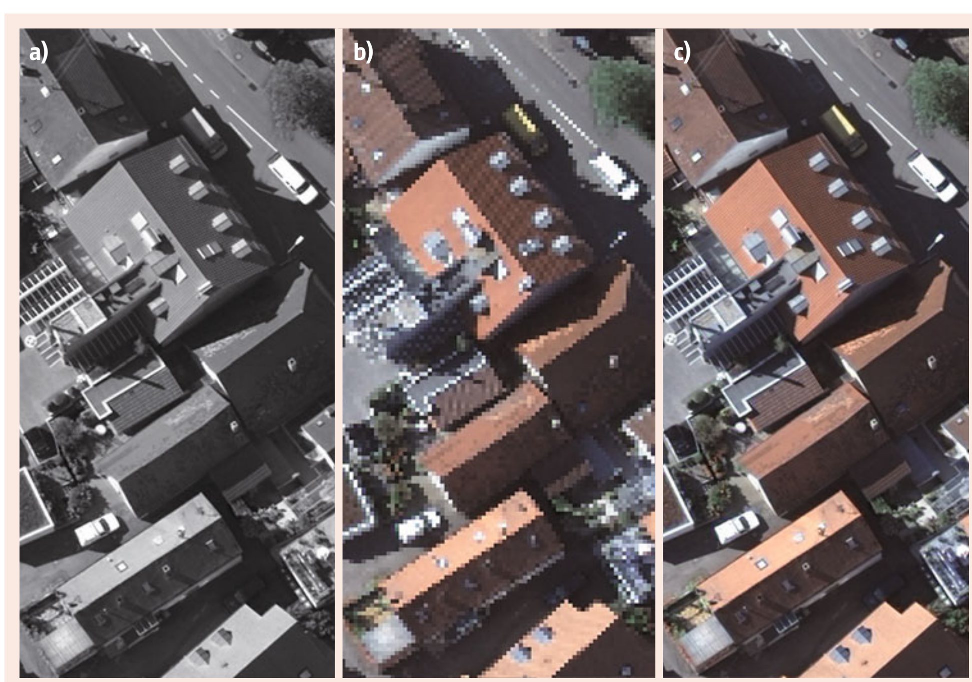
sitive to red and blue, respectively. This is done to adapt to the color sensitivity of human eyes. Full color information is then derived from color interpolation of the neighboring pixels. If, alternatively, several CCD frames are available with each of those being sensitive for one color, the radiometric information can be delivered for each pixel with same resolution. Each pixel on the ground is imaged in all three color bands. Still, the three individual images have to be merged before the full-color RGB image is derived. This approach typically demands at least one camera head per color band or, alternatively, a beam splitter in the camera optic to separate the different color bands within one single camera head.

Color Resolution

Multiple-head cameras usually separate registration of panchromatic (gray values) and color channels. As color is often generated using the Bayer pattern, this results in a somewhat lower geometric resolution with respect to the panchromatic (PAN) image. Also, the design of multispectral channels is likely to set even lower spatial resolution compared to the large-format virtual pan image. High-resolution color imagery is then obtained from postprocessing, where the lower-resolution color channels are combined with the high-resolution PAN images. This process is called pansharpening and is frequently used in satellite imaging. The ratio between the spatial resolution of the pan and color channels is called the pansharpening ratio. Different approaches are used for pansharpening [5]. The methods can be classified into substitution approaches and arithmetic and filter-based techniques. The preservation of the original radiometric color information, depending on the algorithm is exemplarily discussed in [6]. The basic idea of the pansharpening concept is illustrated in Fig. 9.6. This example is taken from digital airborne image data. After the fusion of the lower-resolution RGB image (Fig. 9.6b) with the higher-resolution PAN image (Fig. 9.6a), pansharpening delivers a color image with higher geometric resolution of the PAN channel (Fig. 9.6c). In this example, one low-resolution RGB pixel corresponds to 4×4 high-resolution PAN pixels, which equals a 1 : 4 pansharpening ratio.

9

Fig. 9.6 The concept of pansharpening to increase the spatial resolution of color imagery. (**a**) High-resolution PAN, (**b**) RGB low resolution, and (**c**) RGB after pansharpening. Courtesy of ifp-Stuttgart



9.1.2 Line Sensors

Concept

The previously described group of digital mapping cameras was based on the CCD (or CMOS)-frame concept. Digital imaging from moving platforms might also be based on single or multiple CCD-lines. Similarly to an office scanner, only one or a few CCD lines are arranged perpendicular to the principal moving direction of the sensor. A full 2-D image is indirectly obtained due to the sensor's motion. While the platform is moving, the two-dimensional image data are captured, with the CCD line(s) recording almost continuously. This line scanner concept is also called pushbroom scanning. Digital pushbroom scanners were first introduced into satellite imaging, later also to airborne image acquisition. The principal advantages of a pushbroom scanner is the possibility of extending the length of CCD lines beyond the limits of frame sensors, thus obtaining larger swath and ground coverage. In the modern airborne imaging this advantage is challenged by the introduction of previously discussed virtual frames.

If only one CCD line is used, then the line image has an extension of just one pixel in flight direction. Such a line image is acquired at one distinct point of location and time. The image width equals to the number of pixels per line, i.e., the length of the CCD line. The consecutively imaged lines form the *image strip*, which also is named *image scene*. Notice that each individual line image has its own exterior orientation elements, i.e., position and attitude. This is relevant for the later orientation process of the pushbroom image data (Sect. 9.4). The pixel size obtained on the ground depends on the sampling time of the system and the speed of the platform. Since the linear sizes of the ground pixel in-track and across-track are independent, quadratic pixels on the ground are only obtained if the so-called pushbroom condition is fulfilled. The corresponding ground sampling distance (GSD) is derived from the relation

$$\begin{aligned} \mathrm{GSD}_{\mathrm{along}} &= v\Delta t \\ \mathrm{GSD}_{\mathrm{across}} &= \Delta y m_b = \Delta y \frac{h_{\mathrm{g}}}{f}\,, \end{aligned} \tag{9.1}$$

where Δt is the sampling time, v the speed of the platform, h_{g} the flying height above ground, f the camera focal length, m_b the image scale, and Δy the pixel size across the flight direction.

Typically, more than one CCD line is used in a line scanner system (Fig. 9.7). If two or more CCDs are arranged in one focal plane, along-track stereo viewing becomes possible, where the desired stereo angle is constant and exactly defined through the distance between the different lines in the focal plane. Multiple CCD lines are also necessary to record different color channels. Different to the frame-based approaches no additional sensor heads are necessary for color and multispectral imaging. Additional lines are simply placed in the same focal plane that is already used for the panchromatic channels. Often, at least three panchromatic channels, as well as four multispectral channels, are used. All CCD lines provide the same number of pixels. Thus, panchromatic and multispectral images are obtained with the same geometric resolution, which is again a characteristic different from those of frame-based sensors. Even though almost all systems have more than three lines, such pushbroom systems are often referred to as three-line scanners. This is named after the three panchromatic lines.

Geometrical Configurations

Since the physical location of each of the lines on the focal plane is different, each CCD line provides a different viewing direction, which allows multiple stereo angles within one flight line. In the Fig. 9.7 three multispectral channels are exemplarily placed in the nadir-viewing direction plus

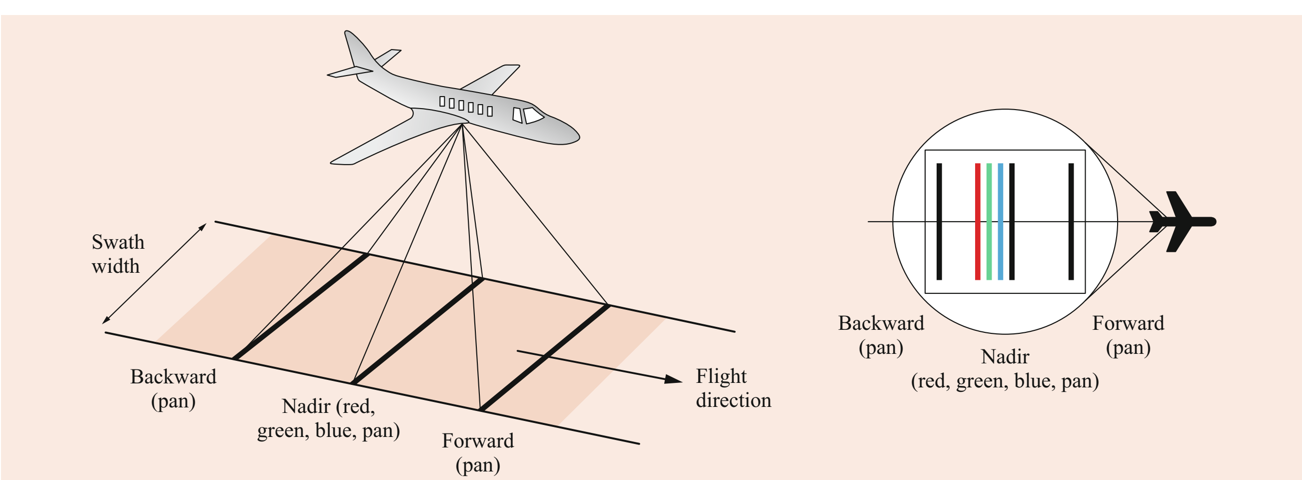


Fig. 9.7 The concept of an airborne line scanning system, after [7]

the three additional panchromatic lines in forward, backward, and nadir directions. Thus, three different stereo angles are possible, namely between the forward and nadir, backward and nadir, and forward and backward views. The color lines might also be arranged in off-nadir direction. In some systems, more than one linear CCD is used for each color channel. If those are placed at different positions in the focal plane this also allows color and multispectrum stereo viewing capabilities.

As mentioned, the stereo capability depends on the different viewing directions due to the parallax effect. Nevertheless, such parallaxes also appear in the different multispectrum lines, which are needed to later obtain the full color images by combination of individual spectral bands. Even though these color/multispectral lines are typically mounted as closely as possible, their displacement will cause different perspective distortions in each spectral band. The larger the distance between the different spectral CCD lines, the larger the influence of these displacements. In order to correct for these effects two options are possible:

- The first is that the full color image is always generated in the orthophoto domain (Sect. 9.5.2). The orthophoto processing corrects for any displacements in the perspective images, also considering the influence of height variations of the imaged scene. If each color band image is fully rectified, the individual bands can easily be overlaid to obtain the full color image.
- Alternatively, this problem can be overcome if so-called beam splitters are installed in the optical system of the pushbroom scanner. Such an installation allows us to exactly coregister the four different color bands (Fig. 9.8). Thus, each of the color bands has the same perspective geometry. Such beam splitters are located in the optic module of the camera, between the lens and the CCDs.

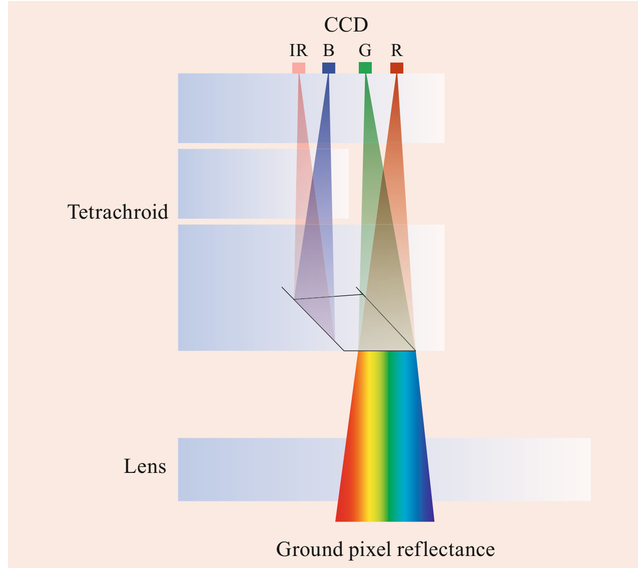


Fig. 9.8 Beam splitter used in a pushbroom sensor (© Leica Geosystems)

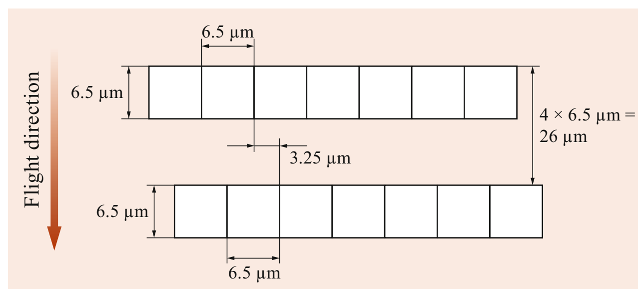


Fig. 9.9 Staggered line arrangement, situation in the focal plane

Image Staggering

The number of pixels per line directly defines the obtainable maximum swath width of the system. If, for example, a GSD of 10 cm is requested, the resulting swath will be 1200 m if the system is based on 12 000 pixels per line CCDs. A larger strip width improves the efficiency of the data capture, as this influences the number of strips to be flown to image a project area.

The width of the swath can be further extended if the image lines are staggered. The staggering means that two CCD lines, so-called A and B lines, are fixed at almost the same position on the focal plane but shifted by half a pixel in the across-track direction. Fig. 9.9 shows the arrangement of a staggered line with $6.5 \times 6.5\,\mu\text{m}^2$ pixel size. Here, the distance between the two lines equals to 4 pixels. While both lines are imaging the same scene, their respective pixel centers are shifted by half a pixel. This obviously increases the sampling interval across flight direction by a factor of 2. The line frequency, i.e., the sampling rate in flight direction, is then adapted according to the new sampling rate in the direction of the CCD line. The A and B lines take one image each, which can be superimposed and combined to a new image with a (nominal) doubled resolution compared to the original images.

Figure 9.10 illustrates the concept of a staggered array. Here, two CCD lines with only 3 pixels per line are combined. It can be seen that when employing the staggered mode, each line acquires pixels of rectangular shape. The sampling rate in the flight direction is duplicated in order to prepare for the later staggering where the form of staggered pixels becomes square. Due to the small distance between the two lines, slightly different parts of the area flown over are imaged at the same time. In this example, line A at time t_2 covers the same area that was already imaged in line B at time t_0. Due to the half a pixel shift between line A and B in the focal plane, the two sampling patterns of both scenes on the ground ideally complement each other, which will deliver

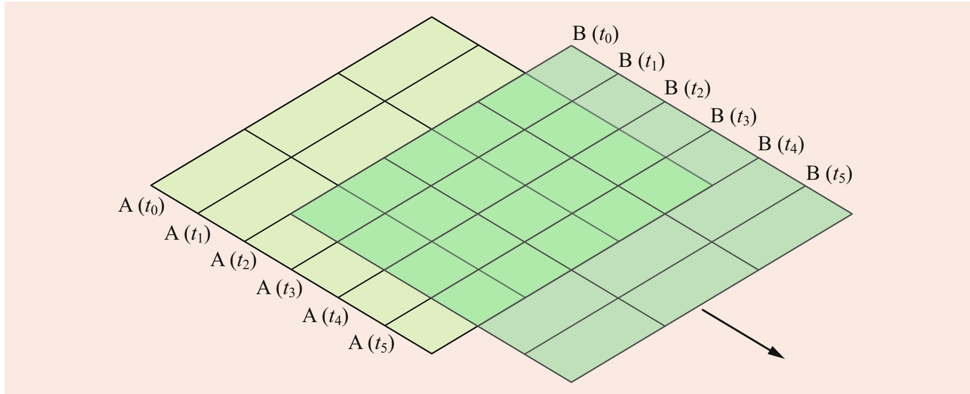


Fig. 9.10 Sampling pattern on the ground with the concept of line staggering. Courtesy of ifp-Stuttgart

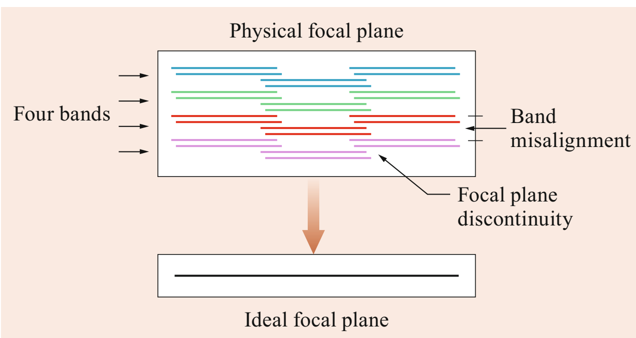


Fig. 9.11 Stepped arrangement of multiple CCD lines

a combined product with increased resolution. The figure also shows that the refinement of resolution fails, if the requested ideal sampling pattern overlap is not done correctly. As lines A and B are physically shifted, there is a small time difference between their exposures to capture the same object on the ground. Therefore, the staggering is affected by the relative change in the sensor attitude during data acquisition and hence the quality of platform stabilization. This effect is less critical for satellite-borne sensors, where the trajectory is much smoother compared to airborne platforms [8].

Another approach to increase resolution and swath width is the employment of multiple CCD lines that are shifted against each other across the flight direction. Since these CCD lines cannot be stitched together directly, a stepped arrangement is necessary (Fig. 9.11). Although this approach has so far been adopted only in the early stages of line-camera development, it is used for obtaining high-resolution satellite imagery. Since the focal plane layout is complex in this configuration, additional processing is required to overcome the discontinuities and misalignment between the lines.

9.1.3 Lidars

The acquisition of terrain structure is very efficiently achieved by optical sensors such as radar and lidar. Both methods are active, i.e., an energy is emitted from the sensors, and its reflection by the object is recorded and processed. The terrain models of highest precision and resolution are usually obtained by airborne laser scanning (ALS), which is in the primary focus of this section.

Laser Ranging

Introduced towards the end of the last millennium, lidar is one of the most important geospatial data acquisition technologies. Together with the state-of-the art navigation technology mobile lidar systems are capable to collect three-dimensional data in large volumes, high density, and with unprecedented accuracy.

The fundamental principle of laser ranging is the ability to measure the travel time t of an emitted laser pulse along its path from the instrument to the target and back (Fig. 9.12). Hence, the distance ρ from the ranging unit towards the target is deduced by the following relation

$$\rho = \frac{1}{2} c\, t \,, \tag{9.2}$$

where c is the speed of light. As shown in Fig. 9.12, the laser-ranging unit comprises an emitting laser and an electro-

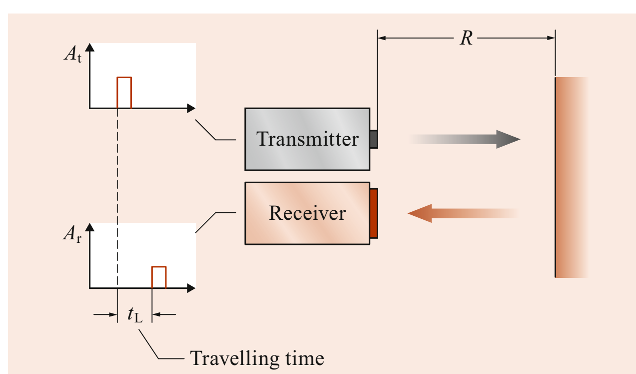


Fig. 9.12 Lidar ranging principle; A_t and A_r are the amplitudes or intensity of the transmitted and received pulses, respectively, after [9]. Courtesy of ISPRS

optical receiver. The transmitting and receiving apertures are oriented in the same direction to ensure that the system will detect the target to which the transmitter points. The size of the laser footprint is a function of the distance to the target and the divergence ϵ of the beam. The angle ϵ defines the instantaneous field of view (IFOV). The IFOV usually spans 0.1–3 mrad.

There are two technological principles of laser ranging that are implemented in mapping applications: continuous wave (CW) lasers and pulse lasers. In CW lasers, the radiation is emitted as a continuous beam instead of a sequence of discrete pulses. This limits the power of the CW laser to terrestrial laser scanning, although there are exceptions [10]. CW lasers deduce the range by comparing the phases between the outgoing and incoming signals. The phase difference of the received light wave is proportional to the travel time of one wavelength (period) and, thus, to the range

$$t = \frac{\phi}{2\pi} T + nT \,, \qquad (9.3)$$

where t is the total elapsed time, ϕ the phase difference of the returned wave, T the period of the modulated signal, and n the number of full wavelengths included in the distance from the transmitter to the receiver. As the phase information is ambiguous for a single measurement, the CW instrument needs to employ a way for its resolution. This can be achieved by various means, most often by modulation of frequency or by following range changes [9].

Although CW lasers reach higher-ranging accuracies, today's airborne lidar systems almost solely use pulsed ranging. A pulse laser functions as shown in Fig. 9.12. The widespread of pulse lasers is due to two technological advances. First, the progress in accurate quartz-stabilized oscillators enables determining the time elapsed between the emission and the reception at picosecond (ps) level (i.e., 10^{-12} s); second, the existence of powerful laser sources with fast shutter limits pulse duration below nanosecond (ns) (i.e., $< 10^{-9}$ s) level. Today's pulse lasers achieve cm to mm-level ranging resolution in long and close-range instruments, respectively [11]. In long-range (airborne) applications, the different implementations of pulse-based rangefinders can be distinguished:

- *Linear mode–discrete echo*. After emission of a high-energy, longer laser pulse, a representative trigger signal of a return (an echo) is detected in real time using analog signal processing. As a discrete pulse is spread in space along its line of sight; part of its energy can be reflected by multiple targets. This allows us to scan even through the canopy, because the spacing between the leaves and branches allows parts of the pulse to penetrate further into the ground, while some energy is reflected immediately. This principle is schematically depicted in Fig. 9.13. As shown in the third plot of Fig. 9.14, the partial reflections are detectable above a certain threshold as distinct peaks in the gathered return signal. These are then discretized into separate echoes. Systems based on this principle can record several returns with minimum separation between successive pulses of several decimeters.
- *Linear mode–full-waveform*. Employing also high-energy, longer laser pulse, these instruments digitize the entire analog echo waveform, i.e., the time-dependent variation of received signal power, for each emitted laser pulse (lowest plot in Fig. 9.14). This approach overcomes the pulse-separation limit present in discrete echo systems and allows finer resolution in the range. The digitization is performed typically on several channels with an inter-

Fig. 9.13 Principle of multiple echoes from a laser signal. After [12], courtesy of the author

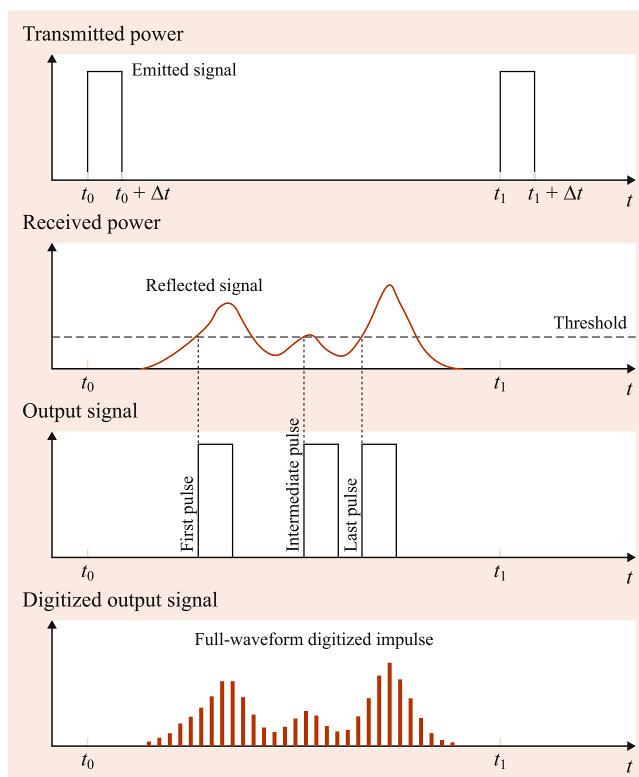


Fig. 9.14 Emitted and received impulse for discrete echo scanners and full-waveform scanners. After [12]

val of 1 ns, which corresponds to spatial quantization of about 0.15 m. The determination of the individual echoes is usually performed after the mission, although modern airborne laser scanning (ALS) systems perform full echo digitization and waveform analysis in real time.

- *Geiger mode*. These devices emit medium energy of short laser pulses into one beam of certain opening. The detector side contains several thousands of pixels that are sensitive to a weak return (few photons) in a binary manner. This so-called "Geiger" counter requires a hundred times less power to register a return than linear mode lidar detectors. The large sensitivity of Geiger detectors allows considerably longer ranging than that of linear mode scanners. Coupling long-ranging capability with the employment of a large number of small detectors and a high repetition rate (hundreds of MHz) allows maintaining a few pulses per m^2 from 5 to 10 km above ground, which increases the swath width and, thus, the productivity considerably. However, the first generation of these detectors allows us to register one (first) echo only with considerably lower precision than that of linear mode scanners. These instruments are yet to be introduced into civilian airborne laser scanning.
- *Single-photon*. These devices emit very low energy and short laser pulses into approximately hundreds of beams. There are separate detectors per beam containing on the order of hundreds of pixels. Each pixel can detect single-photon return at high resolution (< 0.1 m) while registering multiple returns per laser shot with a separation of 1–3 ns. As the system is able to record multiple events per pixel channel and per laser shot in one beam while employing multiple beams, several million points per second are scanned with multiple stops. This technology is, therefore, even more productive than Geiger-mode scanners, albeit not yet as precise as linear-mode lasers. The first commercially available ALS of this type was introduced in 2018. In the same year, a single-photon scanner was placed on an orbit of a satellite mapping ice (ICESat-2).

Most commercial laser rangers operate between 900 and 1500 nm (near-infrared) wavelength, while single-photon lasers currently use 530 nm (green lasers). The amplitude of the backscattered energy A_r is in practice referred to as intensity and is recorded together with the distance observation. (This reference is common but incorrect due to adaptive amplification of the received signal according to its long term average.) Its value depends on several factors:

- *Laser wavelength and target reflectance*. Varying the laser wavelength results in different reflectance responses on the same surface. For example, a laser using wavelength ≈ 1500 nm has good reflectance responses on dark surfaces and manmade structures, whereas surfaces with water content (i.e., glaciers, snow) reflect weakly. On the other hand, systems with shorter wavelength (< 1000 nm) have good reflectance on snow cover but are less optimal for mapping in urban areas. At the same time, objects with high reflectivity, such as street mark paintings or cement, contrast distinctly with objects of low reflectivity, such as coal or soil.
- *Incidence angle of laser beam*. The level of the backscattered signal is a function of the integrated energy distribution across the whole footprint. Accordingly, the larger the incidence angle, the larger the footprint and, consequently, the smaller the backscattered energy.
- *Atmospheric illumination and attenuation*. External illumination, such as sunlight or reflectance from clouds acts as noise in the returned signal. Additionally, light propagation in the troposphere is affected by both scattering and absorption characteristics of the atmospheric medium, thus reducing the reflected energy.

Profilers

Laser profilers measure the distances to a series of closely spaced points distributed along a line on a terrain. In space or airborne applications the profiler is a simple laser ranger (often called laser altimeter) that is pointed towards the ground. Such altimeter measures the distances while is moved over the ground on board a vehicle. As is schematically shown in Fig. 9.15, the 2-D terrain profile is obtained when the altimetric distances are connected to the position and orientation of the laser profiler. Before the invention of satellite positioning, the precise measurement of a carrier's position was difficult to achieve, which was the reason that laser altimetry was used almost exclusively on spaceborne platforms. There, the motion was determined by satellite-tracker observations and by

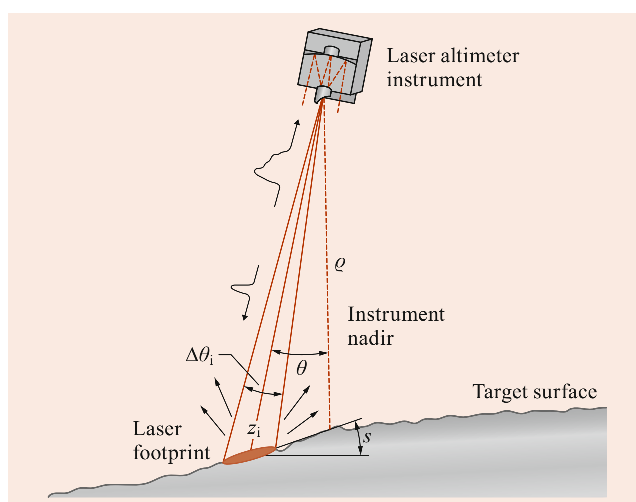


Fig. 9.15 Lidar profiling from a spaceborne platform using a laser altimeter

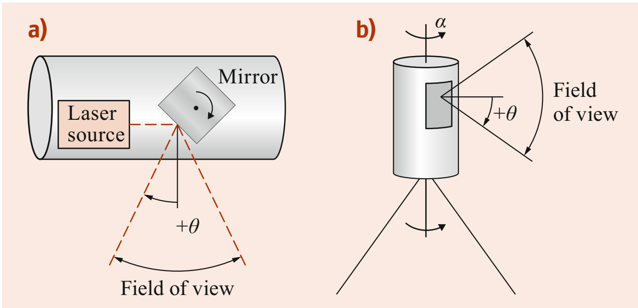


Fig. 9.16 Lidar scanners: (**a**) 2-D laser scanner with a rotating polygonal mirror, (**b**) 3-D scanner based on the same principle

appropriate modeling of the trajectories. This laser technology was first used to determine sea-surface topography, ice cover, desert topography, etc. (e.g. TOPEX/Poseidon, Jason-1, Envisat satellite missions). Later, more sophisticated laser instrumentation allowed the conjoint observation of the Earth surface relief and vegetation canopies (Shuttle Laser Altimeter (SLA), [13]) or distribution of clouds and aerosol (Geoscience Laser Altimeter (GLAS)).

Airborne laser profilers are less common than laser scanners. Nevertheless, these instruments are still used for surveying slowly changing surfaces such as ice-covered terrain [14], lakes or costal water bodies. The latter applications are often connected to the calibration of satellite altimeters or to the study of local gravity field [15].

In a terrestrial or ground-based laser-profiler, a sequence of distance measurements is executed in a series of steps with the slight change of laser-beam orientation between them. Thus, the 2-D elevation profile Δh with respect to the leveled instrument is obtained as

$$\Delta h = \rho \sin(\theta) \qquad (9.4)$$

where ρ is the slant distance and the θ the recorded vertical angle. This results in a two-dimensional profile or vertical cross section of the ground. The terrestrial laser profiler is essentially a 2-D laser scanner which is described in the following section (Sect. 9.1.3, *Scanners*).

Scanners

Lasers scanners combine a laser range-finder with a scanning mechanism (e.g. a mirror) to direct the laser beam into desired direction. The scanning mechanism has either one or two degrees of freedoms that are used to create 2-D or 3-D profiles, respectively. Frequently, the 2-D scanning mechanism is used (Fig. 9.16a), from which the 3-D profile is created by either:

- Rotating the whole scanner assembly along a vertical axis, as would be the case in static Terrestrial Laser Scanning (TLS) (Fig. 9.16b)
- Movement of the carrier in kinematic laser scanning (airborne or vehicle-based scanning).

Thus, in the latter case, the motion of the platform enables along-track scanning, while the mirror deflection provides across-track scanning. The total across-track scanning angle defines the swath width or scanner's field of view (FOV). The swath width SW on the ground can, therefore, be computed as a function of the flying height h and the instrument's FOV ϕ_{max} as

$$\mathrm{SW} = 2h \tan \frac{\phi_{max}}{2} \,. \qquad (9.5)$$

The typical FOV of today's scanners is 50°–60° in airborne and 80°–180° in terrestrial scanning. Several scanning mechanisms exist. The principle of several scanning principles used on airborne platforms is depicted in Fig. 9.17 and their comparison is provided in Table 9.1.

The potential of employing laser ranging for navigation and collision avoidance systems initiated the development of devices operating over shorter distances (< 100 m) without the scanning mechanism. There, a few tens of lasers are arranged in a line array with a regular angular separation and

Fig. 9.17 Different scanning patterns (after [16]): (**a**) polygonal mirror, (**b**) oscillating mirror, (**c**) fiber scan, and (**d**) Palmer scan

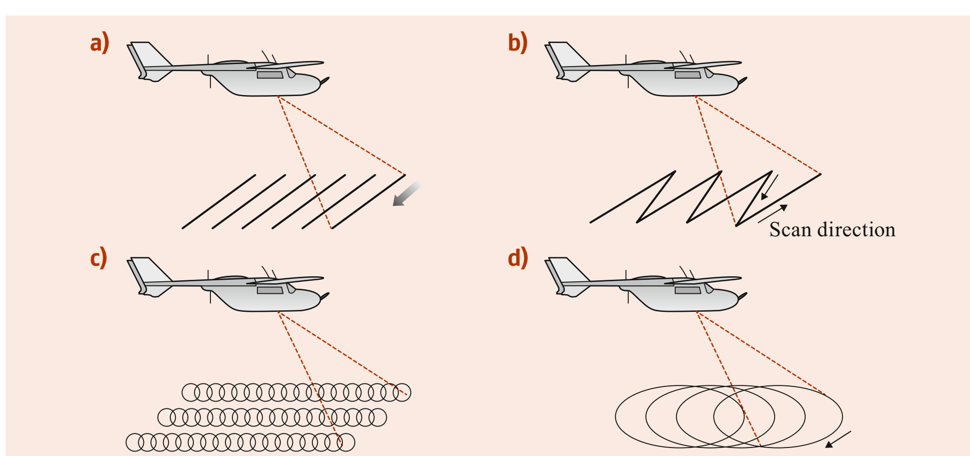


Table 9.1 Comparison of different scanning patterns used in mobile laser scanning

Mechanism	Characteristics	
	Advantages	Disadvantages
Polygonal mirror	• Constant rotation avoids mirror distortions due to additional force • Provides regularly spaced sampling along and across track	• Observations can be taken only at small portion of each mirror facet • FOV is fixed and cannot be adapted • Systems are limited to lower flying heights above ground ($<$ 1000 m)
Oscillating mirror	• Continuous data acquisition possible as mirror points always towards ground • Possibility to compensate aircraft rotation around roll • FOV can be adjusted	• Mirror acceleration causes systematic distortions due to torsion • Z-shaped irregular sampling with lower density at nadir
Fiber scan	• High scan rate possible due to fewer and smaller moving parts • Scan rate sufficiently high to provide along-track overlap • Regular ground sampling	• FOV is limited • Across-track spacing is fixed
Palmer scan	• Scanning is performed twice, each time from a slightly different perspective • Scan rate sufficiently high to provide along-track overlap	• Increased complexity of two mirror motion is harder to calibrate and encode • FOV is limited • Across-track spacing is fixed
Line array	• Faster than a scan due to concurrent use of many lasers • 3-D scan is created / updated rapidly	• Limited to close-ranging with lower accuracy • FOV is limited • Across-track spacing is limited to the number of lasers

FOV of 30°. The 3-D profile is created by rotating the whole assembly, similarly to in Fig. 9.16b, nevertheless with rotation rates up to several tens of Hz, resulting in a high data collection rate. Although the ranging is generally less precise than for scanning lasers, after proper calibration [17, 18], these devices have applicability in mapping from ground vehicles and UAVs.

Scanner frames

The definition of a scanner frame is chosen arbitrarily and, therefore, differs among manufactures. The following definition applies to several systems and can be ported to other instruments by a simple permutation of axes. The location of a point within the scan line j can be conveniently expressed either by polar or Cartesian coordinates, with the former usually being used. Considering the situation depicted in Fig. 9.18, the relations between the *range measurement* ρ, the encoder *horizontal angle* θ, and the *vertical angle* α with respect to the scanner frame defined in Cartesian coordinates are

$$\rho_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2 + z_{ij}^2}\,, \tag{9.6}$$

$$\theta_{ij} = \arctan\left(\frac{y_{ij}}{x_{ij}}\right), \tag{9.7}$$

$$\alpha_{ij} = \arctan\left(\frac{z_{ij}}{\sqrt{x_{ij}^2 + y_{ij}^2}}\right). \tag{9.8}$$

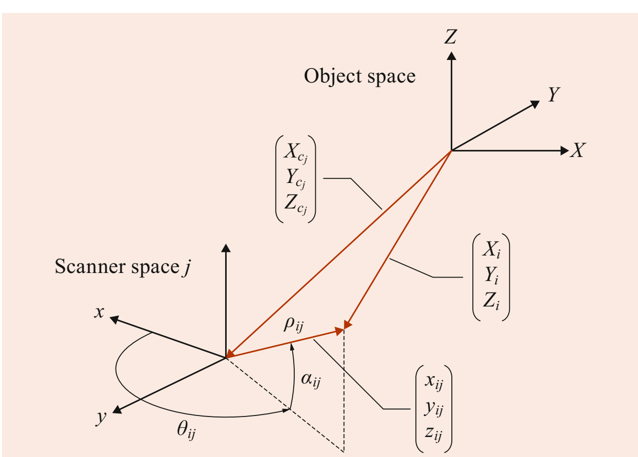


Fig. 9.18 Scanner frame and observation geometry

In the case of a 2-D scanner (e.g. airborne or terrestrial mobile scanning), the angle α is zero and the Cartesian coordinates of the target are expressed as

$$\boldsymbol{x}^s = \rho \begin{pmatrix} 0 \\ \sin\theta \\ \cos\theta \end{pmatrix}. \tag{9.9}$$

9.2 Navigation Sensors

9.2.1 Mapping Prerequisites

Spatial interpretation of remotely sensed data requires determination of the geometric relation between the sensor and the real world. Once these relations have been found, the data can be interpreted in some reference frame (local or global). In the literature, this process is referred to as *georeferencing*,

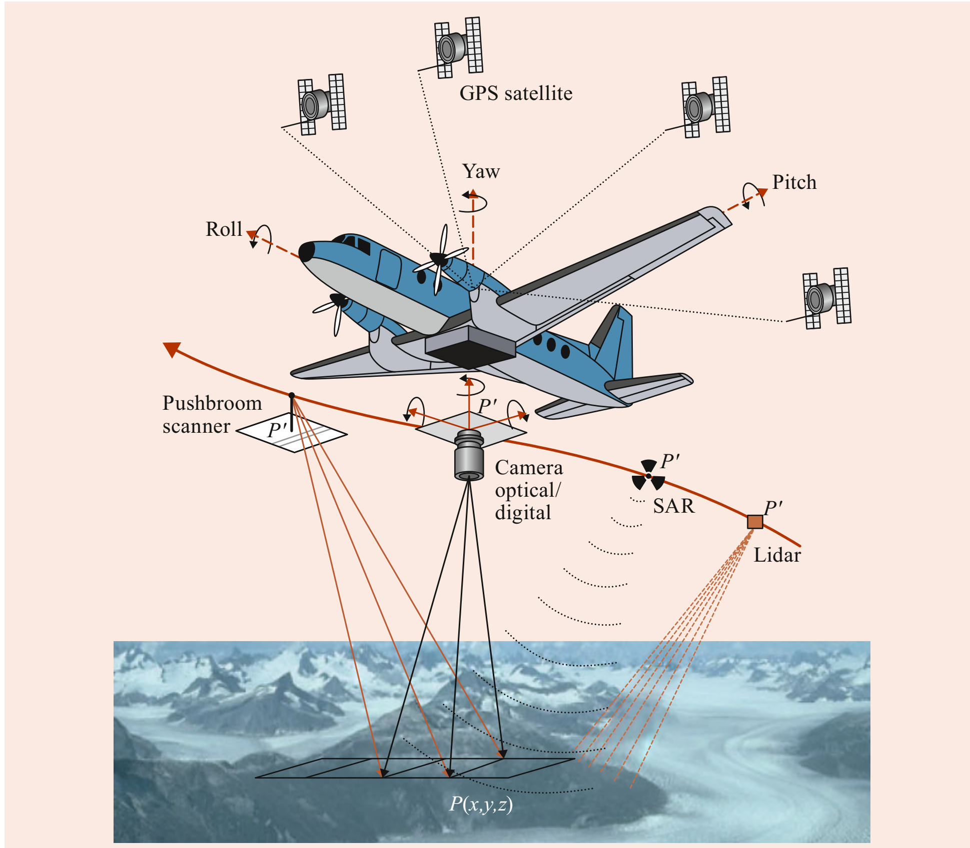


Fig. 9.19 Use of navigation technology for sensor orientation

9

geocoding, or *(sensor) orientation* and concerns the following components (Fig. 9.19):

- The determination of internal geometry of the sensor (*interior orientation*)
- The determination of sensor orientation relatively between scenes (*relative orientation*) or with respect to an external frame (*absolute orientation*).

According to the sensor type the exterior orientation parameters (EO) may include position, attitude (e.g., cameras, scanners), and velocity (e.g., radar). For passive sensors (e.g., frame or line cameras), these parameters may be deduced indirectly from data overlaps and ground control features distributed across the scene (*indirect* sensor orientation) by determining them with a suitable navigation system (*direct* sensor orientation) or by combining both approaches (integrated sensor orientation, Sect. 9.4). Active sensors (e.g., laser, radar), on the other hand, urge the use of direct sensor orientation. Due to the sequential measurement principle and the motion of the carrier vehicle in mobile mapping the EO parameters differ for every object point. The following text provides an overview of the navigation technology that facilitates the problem of sensor orientation tremendously.

Direct measurement of EO parameters typically relies on integrating receivers of the Global Navigation Satellite System (GNSS), such as GPS (Global Positioning System), with an inertial navigation system (INS) with the possible aid of other sensors, whose choice depends on the type of the carrier (e.g., odometers in cars or robots, barometers in aircrafts or drones, star-trackers on satellites). In a GNSS/INS system, GNSS data provides absolute position and velocity information, as well as the error control of inertial measurements, while the INS contributes with attitude estimation, with the interpolation of the trajectory between GNSS position solutions and with the mitigation of sudden perturbations in GNSS measurements (e.g., cycle slips). Both technologies will be introduced separately, while their integration will be described in Sect. 9.2.4). The end of this section is devoted to the introduction of the reference frame and the establishment of relations for transferring the trajectory observation to sensors.

9.2.2 Satellite Navigation

Available Systems

Satellite navigation have global or regional character (Fig. 9.20). There are four global navigation satellite systems

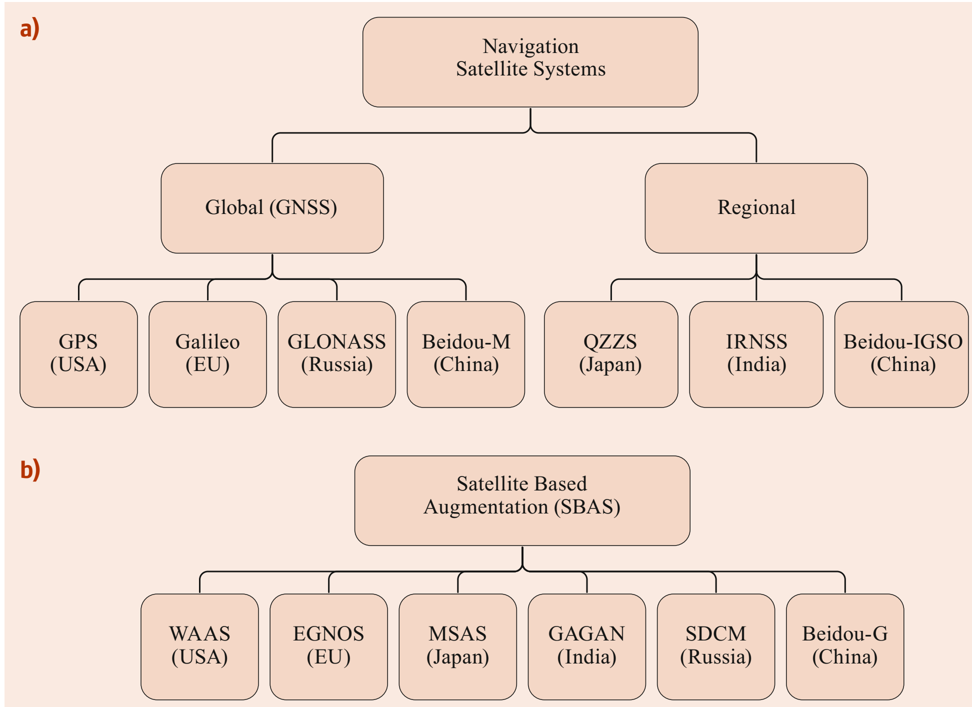


Fig. 9.20 (**a**) Overview of today's GNSS. *Left branch*: satellites on global orbits. *Right branch*: satellites moving only above regions. (**b**) Satellite-based augmentation system with regional implementation of stationary satellites

Table 9.2 GNSS nominal constellation characteristics, after [19]

Characteristics	GNSS			
	GPS	Galileo	GLONASS	Beidou-M
Number of satellites	24–36	24–30	24–30	24–30
Orbital planes	6	3	3	3
Orbital altitude (km)	20,200	23,222	19,130	21,400
Orbital period (hh:mm:ss)	11:58:02	14:04:41	11:15:44	12:52:04

(GNSS), in reverse chronological order, by USA (GPS), Russia (GLONASS), Europe (Galileo), and China (Beidou-M). As of 2020, all systems are fully operational with somewhat similar constellations of 24 to 30 satellites (plus several spares) regularly organized into 6 (GPS) or 3 (others) orbital planes at medium Earth orbit (MEO). The slight differences in orbital altitude among constellations result in different orbital periods, as denoted in Table 9.2. Through regular (GPS) or frequent (GLONASS and Beidou) satellite replacement and late deployment of Galileo, the open radio-navigation satellite service (RNSS) of each constellation uses either identical or very close frequencies and similar signal structures, so the systems are interoperable.

Regional satellite navigation employs Inclined Geosynchronous Orbit (IGSO) with nine (QZZS), seven (IRNSS), and three (Beidou-IGSO) satellites, respectively. In addition to emitting proprietary navigation services, these regional satellites also broadcast open RNSS to enhance GNSS availability over subcontinental areas. Similar enhancement is made by satellite-based augmentation system (SBAS) put forward by civilian aviation authorities. Because it used exclusively geostationary Earth orbits (GEO), SBAS is practically reaching a global coverage (bottom part of Fig. 9.20). Among them WAAS, EGNOS, MSAS, and GAGAN are certified as operational in meeting the exigence of employing GNSS for civilian aircraft navigation in terms of accuracy, integrity, continuity, and availability. Apart from functioning as additional GNSS satellites, SBAS monitors GNSS and provides timely warnings if their signals do not meet the required specifications. Especially for receivers operating on a single-frequency, SBAS significantly improves the accuracy of height determination over the monitored/certified regions.

Signal Structures

The situation of signals on current satellite navigation systems is complex due to evolution. As for GPS, the early satellites broadcasted signals only on two frequencies L1 (centered at 1575.42 MHz) and L2 (centered at 1227.60 MHz), while modernized GPS also includes L5 (centered at 1176.45 MHz). The signal that remains open to all users on all generations of GPS satellites is the coarse-acquisition (C/A) code transmitted on L1. The later generation of GPS satellites emit an additional open signal (L2C) on L2 frequency, while modernized GPS added an open L5 signal on a third frequency and an open L1C sig-

Table 9.3 Open signals of modern GNSS based on CDMA

GNSS	Frequency (MHz)		
	1176–1207	1227	1560–1600
GPS	L5	L2C	C/A, L1C
Galileo	F5 (a+b)		E1
GLONASS	L3OC		L1OC
Beidou	B2 (a+b)		B1-C

nal on the first frequency (Table 9.3). The signal complexity increases from C/A over L2C to L5 and L1C with the goal of improving ranging accuracy, increasing robustness, mitigating adverse effects such as multipath, while improving interoperability with Galileo and other systems.

To distinguish signals coming from different satellites GPS, Galileo and Beidou adopted code division multiple access (CDMA), while GLONASS used frequency separation (FDMA), which made the fabrication of receivers more complex. To improve the interoperability with other systems, modernized GLONASS added CDMA on three frequencies while keeping FDMA for continuity. As shown in Table 9.3, the open service with CDMA on GLONASS is, however, only available on two frequencies. The situation is somewhat similar for Galileo and Beidou, which both adopted complex message structures on E5 and B2 that resulted in a low noise level in code-based ranging (< 0.1 m). The full benefit of all these signals comes to its full potential when broadcasted by a large part of the satellites in every constellation. Due to the interoperability among GNSS the number of available satellites increased substantially over the last decade. In addition, a combined single-frequency GPS/Galileo/GLONASS/Beidou receiver is not significantly more expensive to manufacture than for one system. As explained further, receivers accessing signals on additional frequencies further improve the accuracy and reliability of satellite-based positioning.

Positioning Methods

An overview of current GNSS positioning techniques is provided in Fig. 9.21. The selection of a particular method depends on the factors of accuracy and rapidity in data acquisition and mobile mapping. These methods are the following:

- *Single-Point Positioning (SPP)* is the most commonly used method for real-time positioning. It is based on a single receiver and phase-smoothed code data processing (absolute GNSS positioning). Provided that SBAS corrections for altitude are available, this approach can deliver accuracies of 0.5–3 m. However, SBAS is not available worldwide, and the reception of the ionospheric correction grid emitted by the geostationary satellite (which orbits above the equator) is valid only inside the monitored region. In such a case, the *Standard Positioning Service (SPS)* provides an accuracy of about 2 to 10 m.

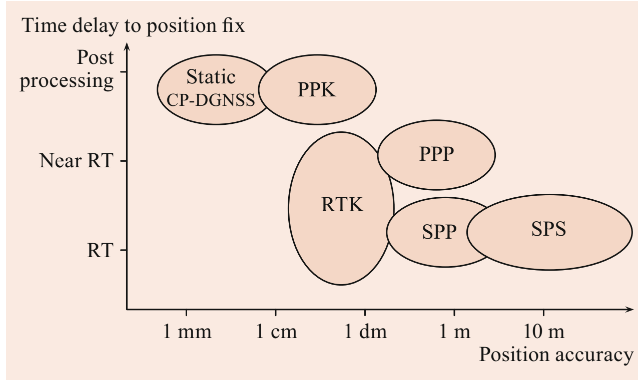


Fig. 9.21 Overview of GNSS positioning methods as a function of accuracy and rapidity

- *Precise-Point Positioning (PPP)* is a novel positioning methodology based on the fast availability (i.e., within an hour) of precise GNSS satellite orbit parameters and clock corrections. This technique can achieve subdecimetric position accuracy [20] and is available worldwide without the need for an augmentation system.
- *Differential GNSS (DGNSS)*, *carrier-phase DGNSS* and *postprocessed kinematic (PPK)* are relative positioning techniques based on simultaneous observations by the rover and base (one or more) receivers, where the latter is placed at a location with known coordinates. DGNSS uses only the code (or carrier-smoothed code) observations, while the other two also employ the more precise but ambiguous carrier-phase measurements. The ambiguities are resolved via complex processing, whose reliability is increased with dual-frequency observations. For *static* carrier-phase DGNSS, subcentimeter to millimeter accuracy can be achieved when respecting some considerations about baseline length and observation time. The upper limit in PPK is centimeter to decimeter accuracy for a relative baseline length limit of around 15 km when using carrier-phase observations on two (or more) frequencies. The accuracy of PPK without ambiguity resolution is generally not better than a few decimeters, while ambiguity resolution using observations on a single frequency is limited to a 1–2 km long baseline.
- *Real-time kinematics (RTK)* applies the above-mentioned PPK principles in real time. Its prerequisite is the establishment of a communication link transmitting reference measurements or correction parameters. Similarly to PPK, this information is provided from a base receiver or from a network of receivers. National-wide networks broadcasting such types of corrections are available in many regions and can be accessed via modern communication technologies (i.e., Internet and mobile telephone networks). This makes them employable even for kinematic data acquisition. Subdecimeter to centimeter-level positioning accuracy can be achieved by these means in ideal conditions.

9.2.3 Inertial Navigation

Inertial navigation derives position, velocity, and attitude from the initial knowledge of these quantities and from the integration of the accelerations observed (more precisely, *specific forces*) and *angular velocities* along their motion. These observations are normally obtained from a minimum of three accelerometers and three gyroscopes that are orthogonally mounted in an inertial measurement unit (IMU). An IMU coupled with a navigation computer creates an inertial navigation system (INS). A detailed overview of gyroscope and accelerometer technology can be found in [21, 22]; the following list is limited to the types common in direct sensor orientation.

Gyroscope Technology

Gyroscopes usually represent the most expensive part of an IMU. Their accuracy significantly affects the overall navigation performance of an INS. Several types of gyros are used in sensor orientation:

- *Mechanical gyros*. These gyroscopes use the principle of conservation of the angular momentum. A mass is spun at high speed around its axis, and the reaction to external forces (called precession forces) acting on its spin due to the casing rotation is measured. The most common rotational gyro employed for sensor orientation is the *dynamically tuned gyro (DTG)*. It is relatively small, affordable, and provides excellent short-term accuracy.
- *Optical gyros*. Such gyros use the Sagnac effect that rises due to the fact that speed of light is conserved in rotating systems [23]. The most common types are *ring laser gyros (RLG)* and *fiber optical gyros (FOG)*. Both are used for the most accuracy-demanding applications. FOGs of lower category are employed in the wider context of sensor orientation.
- *Vibratory gyros*. These gyroscopes exploit the principle that an oscillating body preserves the plane of vibration in inertial space despite rotations. These sensors are usually less precise, but they are smaller and cheaper to fabricate. They are often employed in airborne applications with middle to low accuracy requirements.
- *MEMS-gyros*. These tiny gyroscopes exploit different physical principles and come in varying sizes and quality through microelectromechanical system (MEMS) technology, which produces small and inexpensive sensors. They are used in mass market, automotive, robotic, and entry-level navigation applications. They are indispensable on drones for control and guidance. As their high end in some aspects approaches the quality of low-end FOGs, they are useful for robotics and UAV-based acquisition.

Accelerometer Technology

For accelerometers, three relevant types have to be mentioned:

- *Force rebalance accelerometers* measure the electrical current that is proportional to the force needed to maintain a suspended proof mass at rest under acceleration. These are used in the most demanding autonomous or airborne applications (e.g., precise underground or indoor mapping of large structures or high-altitude flights).
- *Vibrating accelerometers* exploit the resonant frequency of a mass hanging on a vibrating string. The frequency of vibration varies when an acceleration acts in the direction of the string. Such accelerometers are often fabricated as high-end MEMS sensors.
- *MEMS-gyros* based on different physical principles. Their high-end type is employed in robotics and UAV-based acquisition.

Strapdown INS

In earlier INS realizations, inertial sensors were mounted on stabilized (gimbaled) platforms, thus mechanically isolated from the rotational motion of the carrier. The advances in digital processing made it possible to avoid gimbaled mounts. Nowadays, inertial sensors are rigidly mounted (strapped down) to their casing, thereby decreasing the complexity and cost of the system while increasing the dynamic range of motion that can be tracked. As the number of moving parts is reduced, these systems are also smaller and more reliable. A strapdown INS is often fitted in the same casing together with an optical instrument, and its orientation output can be used for sensor-head stabilization (Sect. 9.4.6).

9.2.4 Integrated Navigation

Principle

Integrated navigation is a technique that combines data from several navigation systems or sensors with the aim of improving the accuracy and robustness of the estimated trajectory. In this respect, satellites and inertial navigation have a very different but complementary behavior. The performance of a standalone INS is characterized by a time-dependent drift in the accuracy of the position, velocity, and attitude estimates it provides. The rate at which the navigation errors grow in time is governed predominantly by the accuracy of the initial alignment, noise, and imperfections in the inertial sensors and their assembly, as well as the dynamics of the trajectory. Whilst improved positioning accuracy can be achieved through the use of more accurate sensors, this cannot match the GNSS-type precision in the long run.

On the other hand, the GNSS positioning is conditioned by the requirement of the line of sight to a number of satellites (four or more), which is difficult to maintain in all situations, especially in terrestrial mobile mapping or indoors. Therefore, the combination of both systems enhances the trajectory determination across the spectrum of motion.

Contrary to GNSS, inertial navigation provides continuous data output for all trajectory parameters (i.e., position, velocity, and attitude). Therefore, the integrated navigation principally stabilizes and refines the INS output by estimating and correcting the systematic effects in the inertial sensors and in the initialization. Different types of navigation aiding may be categorized as follows:

- *External measurements*: measurements obtained by receiving signals or by viewing objects outside the vehicle. Such observations may be provided by radio navigation aids, GNSS satellites, star trackers or imagery, for example.
- *Autonomous measurements*: measurements derived using additional sensors carried on-board without the dependence on external infrastructure or visibility. Navigation of this type may be provided by odometers, pressure sensors, Doppler radar, or magnetic sensors, for example.
- *Dynamic constraints*: the application of implicit knowledge of some dynamical state or its form. For example, constraints such as zero velocity and nonholonomic condition (i.e., the alignment of vehicle speed with its direction) are used as a supplementary aiding method on terrestrial vehicles or complete vehicle dynamic model is used for UAVs [24, 25].

Integration Schemes

Optimal integration of different measurement data with inertial observations is commonly achieved by using a Kalman filter/smoother (Fig. 9.22). Data filtering/smoothing can, however, be organized in different manners with respect to GNSS observations. The following two integration schemes are the most important:

- *Loosely-coupled integration*. This is the most common integration approach, especially in airborne or shipborne installations. Raw IMU measurements are integrated to yield position and attitude at the IMU output rate (normally 100–500 Hz). The position and velocity data gathered by GNSS are processed independently, yielding a sequence of positions and velocities at a certain frequency (normally 0.1–1 Hz). These data are subsequently fed as updates in an extended Kalman filter (EKF). The differences observed between the predicted (INS) and GNSS-determined velocities and positions are used to estimate the elements of the filter state vector, containing, on one hand, the error states related to the trajectory (i.e., position, velocity, and orientation) and, on the other hand, those related to the inertial sensors themselves (i.e., gyro and accelerometers biases and scale factors, odometer or pressure sensor bias, etc.).
- *Closely-coupled integration*. In this integration scheme, GNSS raw measurements (normally double-differenced code, phase, and Doppler measurements) are fed directly into the Kalman filter. Therefore, the GNSS measurements can be used in the filter even if the number of visible satellites is not sufficient to compute an independent position fix (i.e., lower than four). Accordingly, this integration scheme is advantageous for environments with reduced GNSS signal reception (e.g., urban canyons), and is commonly used in terrestrial mobile mapping.

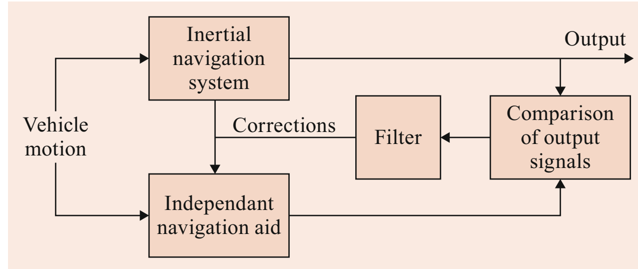


Fig. 9.22 Typical integrated navigation scheme for direct georeferencing

A performance comparison between the two integration schemes presented can be found in [26, 27] and, with a focus on RTK, in [28].

Resulting Accuracy

For strapdown INS, sensor integration solves firstly the problem of calibrating the systematic errors (i.e., residual gyro and accelerometer biases, scale factors, etc.). Secondly, the use of GNSS data mitigates attitude initialization errors and in certain cases, enables kinematic (in-flight alignment), which removes the need for the vehicle to be held stationary for the north-seeking process prior to movement. (This concerns all gyroscopes of lower accuracy, like those employed in UAVs that cannot complete north-seeking without an external assistance.) At the same time, the inertial system smoothens noisy velocity outputs from GNSS and provides high-rate measurement of position and velocity over a larger spectrum of motion.

There is no such thing as a perfect instrument, and as strong as it is, the integration cannot eliminate all errors completely. In other words, the data integration handled by a Kalman filter/smoother only cancels the nonoverlapping part of the sensor's error budget. The performance of error cancelation depends on the motion of interest, the instrument type, and the encountered dynamics. While the long-term positioning accuracy limit depends on the GNSS positioning

Table 9.4 Orientation accuracy of as a function of time and INS quality

Time	Navigation grade		Tactical grade	
	Roll/pitch (°)	Yaw (°)	Roll/pitch (°)	Yaw (°)
1 s	0.001–0.0014	0.001–0.002	0.002–0.02	0.001–0.05
1–3 min	0.0014–0.003	0.004–0.005	0.005–0.04	0.008–0.1
Longer time	Trajectory dependent – similar to 1–3 min when optimal			

Table 9.5 Orientation accuracy for small MEMS IMU of high quality

Time	Low-end tactical grade	
	Roll/pitch (°)	Yaw (°)
1 s	0.005–0.1	0.005–0.2
1–3 min	0.03–0.2	0.1–0.2
Longer time	Trajectory dependent	

solution (Fig. 9.21), the time over which such accuracy can be maintained in the absence of satellite signals mainly depends on the quality of the INS and its preceding calibration. Based on the position error accumulated after 1 h of autonomous operation, the INS are normally grouped into four main categories [29]: *strategic grade*, *navigation grade*, *tactical grade*, and *low-cost* (MEMS) instruments. A summary of potential orientation accuracies for today's most popular sensors used for civilian applications in sensor orientation is summarized in Table 9.4. The automotive in Table 9.5 corresponds to small MEMS IMUs of high quality, like those used by terrestrial and indoor robots or UAVs.

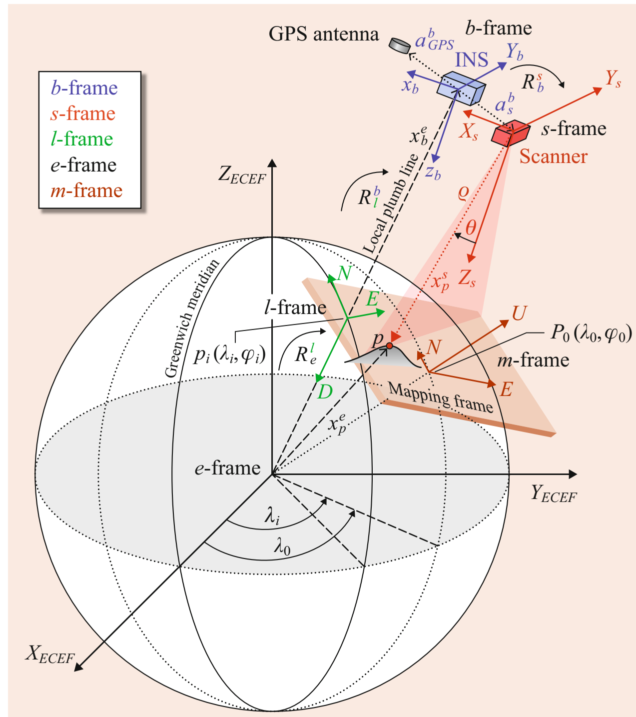


Fig. 9.23 Geometry of direct sensor orientation

9.2.5 Geometrical Relations

Coordinate Frames

Sensor frame – *s*

While the nature of imagery captured in a static environment (e.g., terrestrial laser scanning or terrestrial photogrammetry) allows operations within a local coordinate system, kinematic remote sensing with navigation support requires the employment of a global reference frame, as well as several intermediate frames. Table 9.6 provides an overview of the frames used, together with their abbreviated identifiers. All frames are defined to be right-handed Cartesian frames, and their relation is schematically depicted in Fig. 9.23 for the case of an airborne laser scanner. The situation is similar for other optical devices for which the sensor frames(s) are defined later (Sects. 9.1 and 9.4).

Earth-Centered, Earth-Fixed Frame – *e*

The satellite orbits of common GNSS systems are referred to this frame, and so is the outcome of the trajectory computation. A geocentric ellipsoid is normally attached to an ECEF frame, and its properties together with other geophys-

Table 9.6 Overview of reference frames

ID	Frame name	Description
s	Sensor frame	Frame of the laser sensor, defined by the principal axes of an optical instrument; e.g. xy-axes define an image plane in the frame imagery, yz-defines the scanning plane of a 2-D scanner
b	Body frame	Frame realized by the triad of accelerometers within an IMU
l	Local-level frame	This frame is tangent to the global ellipsoid (normally WGS84), with the orthogonal components usually defined as N-orth (x), E-ast (y) and D-own (z)
e	ECEF frame	Earth-centered Earth-fixed frame. The origin is the geocenter of the Earth, x-axis points towards the Greenwich Meridian, and the z-axis is the mean direction of the Earth's rotational axis. The y-axis completes the right-handed Cartesian system
m	Mapping frame	Cartesian frame with E-ast (x), N-orth (y), and U-p (z) component. The easiest implementation is the local-tangent plane frame, but this frame can also be represented by a projection and/or national datum

ical parameters define a world datum (e.g., WGS84 used for GPS measurements). Coordinates in this frame can either be expressed as *geocentric coordinates* (x^e, y^e, z^e) or as *geographical coordinates* (latitude φ, longitude λ, ellipsoidal height h). The latter parameterization is often used in the output of GNSS/INS trajectory. The relation between the Cartesian and ellipsoidal coordinates reads

$$\boldsymbol{x}^e = \begin{pmatrix} x_1^e \\ x_2^e \\ x_3^e \end{pmatrix} = \begin{pmatrix} (N+h)\cos\varphi\,\cos\lambda \\ (N+h)\cos\varphi\,\sin\lambda \\ \left(\frac{b^2}{a^2}N+h\right)\sin\varphi \end{pmatrix}, \tag{9.10}$$

where N is the radius of curvature in the prime vertical, and a and b are the semimajor and semiminor axes, respectively.

Local-level frame – *l*

This frame is mainly used as the reference for the orientation angle output from GNSS/INS processing. Its origin is defined by the sensor position on a reference ellipsoid at zero height, which corresponds to the intersection of the local vertical at the actual sensor position with the reference surface. The x^l-axis points along the local meridian to the north, the y^l-axis points to the east, and the z^l-axis points downward to complete the system. Such a local frame definition is called l-NED (for north–east–down), while the upward positive convention of the z^l-axis defines the l-ENU frame (east–north–up). The rotation from the l-frame to the e-frame can be described by the matrix $\mathbf{R}^e_{l_{\text{NED}}}$

$$\mathbf{R}^e_{l_{\text{NED}}} = \begin{pmatrix} -\sin\varphi\,\cos\lambda & -\sin\lambda & -\cos\varphi\,\cos\lambda \\ -\sin\varphi\,\sin\lambda & \cos\lambda & -\cos\varphi\,\sin\lambda \\ \cos\varphi & 0 & -\sin\varphi \end{pmatrix}. \tag{9.11}$$

Body frame – *b*

The body frame is represented by the axes of the inertial navigation system. The origin of the b-frame is located at the navigation center of the INS, and the axes are congruent with the axes spanned by the triad of accelerometers. Normally, the b-frame axis coincides with the principal axis of rotation of the carrier or can be rotated to them by some cardinal rotation. According to the aerospace norm ARINC 705, the axis and the rotations describing the 3-D attitude are defined as follows. The x^b-axis points forward along the fuselage, the y^b-axis points to the right, and the z^b-axis points downward. The associated rotation angles along the x-y-z-axes are referred to as *roll* (r), *pitch* (p), and *yaw* (y). Respecting the aerospace attitude definitions, the corresponding rotation matrix that relates the l-frame to the b-frame takes the following form

$$\mathbf{R}^b_{l_{\text{NED}}} = \mathbf{R}_x(r)\,\mathbf{R}_y(p)\,\mathbf{R}_z(y)\,. \tag{9.12}$$

where $\mathbf{R}_x(r)$, $\mathbf{R}_y(p)$, and $\mathbf{R}_z(y)$ are defined as

$$\begin{aligned} \mathbf{R}_x(r) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos r & \sin r \\ 0 & -\sin r & \cos r \end{bmatrix} \\ \mathbf{R}_y(p) &= \begin{bmatrix} \cos p & 0 & -\sin p \\ 0 & 1 & 0 \\ \sin p & 0 & \cos p \end{bmatrix} \\ \mathbf{R}_z(y) &= \begin{bmatrix} \cos y & \sin y & 0 \\ -\sin y & \cos y & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{9.13}$$

Transformation of the Exterior Orientation

The relationship between an arbitrary point $\boldsymbol{x}_p^s$ in the s-frame coordinates and that same vector expressed in the b-frame is given by

$$\boldsymbol{x}_p^b = \boldsymbol{x}_{bs}^b + \mathbf{R}_s^b(\omega_b, \varphi_b, \kappa_b)\boldsymbol{x}_p^s\,, \tag{9.14}$$

where $\boldsymbol{x}_{bs}^b = \mathbf{R}_s^b \boldsymbol{x}_{bs}^s$ denotes the origin of the s-frame in the b-frame, which is also known as the lever-arm vector. The rotation matrix $\mathbf{R}_s^b$ in (9.14) is called the bore sight and represents the relative misalignment between the s and b-frames. This matrix is usually parameterized by the three Euler angles ω_b, φ_b, κ_b. The magnitude of the bore-sight angles and the lever-arm vector need to be determined by calibration.

The observation equation for direct sensor orientation for a point p viewed by a sensor s in the e-frame coordinates follows from Fig. 9.23 by combining (9.10) with (9.12)

$$\boldsymbol{x}_p^e(t) = \boldsymbol{x}_b^e(t) + \mathbf{R}_l^e(t)\mathbf{R}_b^l(t)\mathbf{R}_s^b(\omega_b, \phi_b, \kappa_b)\left(\boldsymbol{x}_{bs}^s + \boldsymbol{x}_p^s(t)\right), \tag{9.15}$$

where $\boldsymbol{x}_b^e(t)$ is the navigation center of the IMU in the e-frame, and all other components were defined previously. The symbol (t) indicates quantities that vary with time.

Mapping frame – *m*

For active sensors as laser scanners, the coordinates of observed points in the ECEF-frame can be generated via (9.15). However, the final coordinates are often needed in some other datum and projection. The so-called mapping frame habitually represents a national coordinate system, and the results of mapping can be transferred to such a frame pointwise or pixelwise via relations published by local surveying authorities. Alternatively, the registration of optical images and that of lasers can be performed directly in mapping frame as discussed in detail in [30, 31].

System Calibration

The method of direct sensor orientation requires that the optical sensor be calibrated for the parameters of interior orientation, which includes system installation. The latter concerns determining the spatial and orientation offsets that exist between optical and navigation sensors.

The lever arm $\boldsymbol{x}^b_{bs}$ is either specified by the system provider or needs to be determined per installation. The same is true for the lever arm between the IMU center and the GNSS antenna $\boldsymbol{x}^b_{ba}$, which is needed during GNSS/INS integration. Calibration of lever arms is best performed by tacheometry. Such a procedure is discussed in detail in [12, 32] for an aircraft and a small UAV system, respectively. An alternative solution is to estimate $\boldsymbol{x}^b_{ba}$ directly in the GNSS/INS Kalman filter/smoother as an additional parameter. Similarly, $\boldsymbol{x}^b_{bs}$ can be estimated in the block adjustment (Sect. 9.4), but its value is often strongly correlated to other parameters, and this approach should, therefore, be avoided when later used for direct sensor orientation.

The recovery of the bore-sight matrix $\mathbf{R}^b_s$ is more involved and requires the use of principles described in Sect. 9.4. For frame cameras, this process can be achieved in either one [33, 34] or two steps [35]. A similar procedure is maintained for line scanners [36, 37]. The bore-sight determination in kinematic laser scanning followed a rapid evolution [38–40] that converged to the approach based on surfaces of known form [41, 42]. The calibration principles are further addressed in Sect. 9.4.

9.3 Photogrammetry

9.3.1 From 2-D to 3-D

The main task of photogrammetry or, equivalently, *computer vision* is to reconstruct 3-D scenes from 2-D images. The most important requirement for the reconstruction to work is the that the scene be imaged from different places, so that sufficient correspondences between pictures can be (automatically) established. Position and orientation of each image is found along the way, which fact allows us to infer the motion of the camera and, thus, the platform (up to the image acquisition rate). At the same time, the knowledge of camera motion observed by other sensor(s), like those discussed in Sect. 9.2, can be used in support of the reconstruction process, which is a subject of Sect. 9.4.

Given a set of images, the principal challenges (and steps) of reconstructing 3-D models are threefold (Fig. 9.24):

- *Correspondences*: automatically detect sufficient number of *key* features on each image and establish their correspondences with other images.
- *Geometry (motion, orientation)*: recover camera pose (position and orientation) between images, its intrinsic parameters (calibration), and features' 3-D coordinates.
- *Scene (structure)*: use the knowledge of geometry to create dense point cloud to recover 3-D objects (models) with texture.

This section essentially concentrates on the geometry part of the reconstruction problem without the help of navigation sensors. It is known under different names, such as *orientation and calibration* in photogrammetry or *structure from motion* in computer vision, as well as *bundle adjustment* (both).

We will describe the process of image formation using a mathematical model that accounts for three types of transformations:

1. Coordinate transformation between image coordinate frames
2. Projection of a 3-D scene onto 2-D image coordinates
3. Relation between the camera frame and an external mapping frame.

This chapter proceeds as follows. Although a general introduction to the relation between sensor-mapping frames was given in Sect. 9.2.5, we repeat it here but in a homogeneous representation of coordinates (Sect. 9.3.2), the form of which will be useful later on. Then, after introducing the basic geometry of the imaging system (Sects. 9.3.3 and 9.3.4), we describe a model of image formation for an ideal perspective camera (Sect. 9.3.5. With the necessary components, we introduce the reconstruction process for a stereo-pair of images (Sect. 9.3.6), which we later extend to multiple views (Sect. 9.3.7). We conclude the chapter with different processing strategies for filtering and optimization in scene reconstruction.

9.3.2 Camera Pose in a Homogeneous Form

Consider two Cartesian frames, where one is a mapping frame spanning the object space, and the other is related to a camera, viewing a scene at certain time t, to which belongs a point p. From the situation depicted in Fig. 9.25, it is clear that the coordinates of a point p with respect to the mapping frame m are simply the sum of the translation $\boldsymbol{x}^m_c$ of the origin of the frame c relative to that of the frame m and the vector $\boldsymbol{x}^c$ expressed in relation to the mapping frame m, which is $\mathbf{R}^m_c\boldsymbol{x}^c$, where $\mathbf{R}^m_c$ is the relative rotation between the frames

$$\boldsymbol{x}^m = \boldsymbol{x}^m_c + \mathbf{R}^m_c\boldsymbol{x}^c \,. \tag{9.16}$$

Every time the camera moves, its motion is captured by $\mathbf{T}^m_c = (\mathbf{R}^m_c, \boldsymbol{x}^m_c)$ or more shortly by $\mathbf{T} = (\mathbf{R}, \boldsymbol{x})$ when the frames involved are clear of the context. It will become an

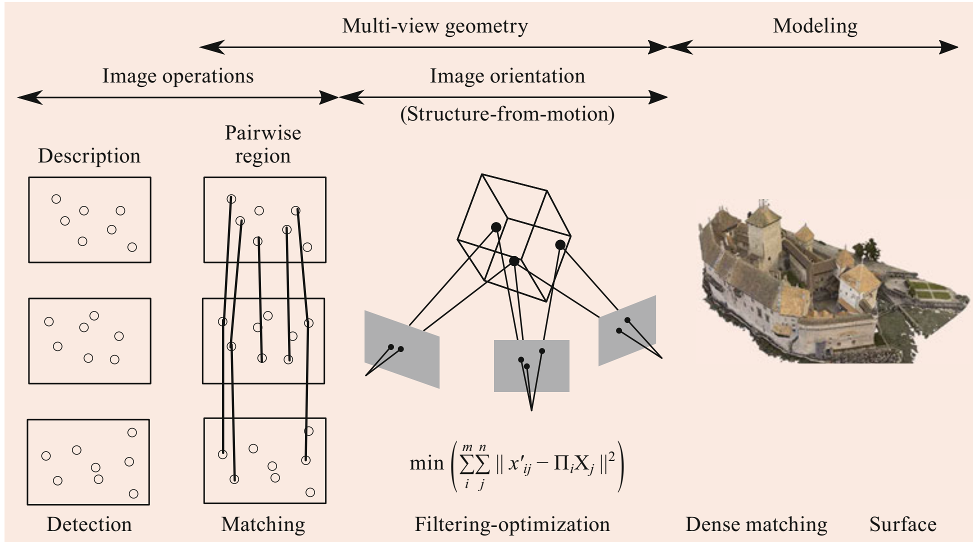


Fig. 9.24 Photogrammetry/computer-vision process for 3-D scene reconstructions based on 2-D imagery

advantage when we convert the transformation expressed by (9.16) to an expression of the form $\boldsymbol{u} = \mathbf{A}\boldsymbol{v}$. This is possible by adding "1" to the vector $\boldsymbol{x}$ as its fourth coordinate and by defining operations on so-called *homogeneous coordinates*. Such an extension preserves the original Euclidean space,

$$\overline{\boldsymbol{x}} \doteq \begin{pmatrix} \boldsymbol{x} \\ 1 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \qquad (9.17)$$

The vectors are defined analogically as differences of coordinates $\overline{v} = \overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2$. Differences makes the fourth component null and give rise to the original subspace. Rewriting (9.16)

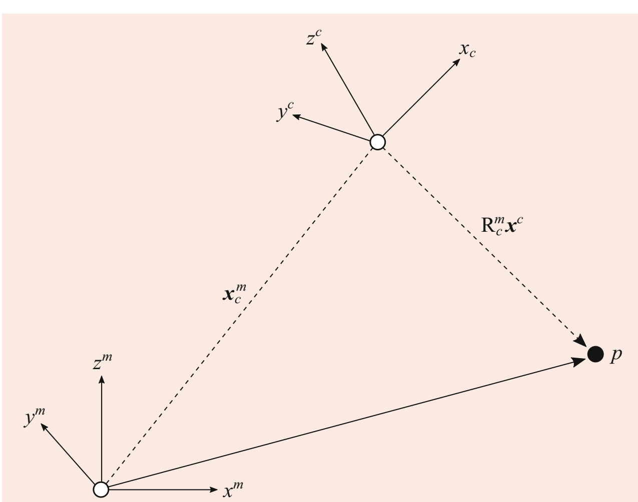


Fig. 9.25 Motion of a camera frame with respect to a Cartesian mapping frame

in the new notation leads to

$$\overline{\boldsymbol{x}}^m = \begin{pmatrix} \boldsymbol{x}^m \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}_c^m & \boldsymbol{x}_c^m \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^c \\ 1 \end{pmatrix} \doteq \overline{\mathbf{T}}_c^m \overline{\boldsymbol{x}}_c , \qquad (9.18)$$

where the 4×4 matrix $\overline{\mathbf{T}}_c^m$ is the *homogeneous representation* of the rigid-body transformation. Now it is possible to encapsulate the coordinate transformation between several frames as a sequence of multiplications,

$$\overline{\mathbf{T}}_c^a = \overline{\mathbf{T}}_b^a \overline{\mathbf{T}}_c^b = \begin{pmatrix} \mathbf{R}_b^a & \boldsymbol{x}_b^a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R}_c^b & \boldsymbol{x}_c^b \\ 0 & 1 \end{pmatrix}. \qquad (9.19)$$

As can be easily verified, the inverse transformation is

$$\overline{\mathbf{T}}^{-1} = \begin{pmatrix} \mathbf{R} & \boldsymbol{x} \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{R}^\mathsf{T} & -\mathbf{R}^\mathsf{T}\boldsymbol{x} \\ 0 & 1 \end{pmatrix}. \qquad (9.20)$$

9.3.3 Pinhole Camera

Consider the basic imaging system as described by Fig. 9.1 in Sect. 9.1; the aperture of the lens decreases to zero, and the only feature that contributes to the illumination of an image point is that on the line going through the center of the lens o. This way, an *image point* can be directly related to an *object point*, as shown in the upper part of Fig. 9.26, where, for simplicity, the camera frame is oriented in the same direction as the object frame, and the point p is such that its image coordinate $y' = 0$.

Let us define the distance from the object point to the optical center along the optical axis as Z, and the "horizontal" projection of the point on the optical axis as X, while Y

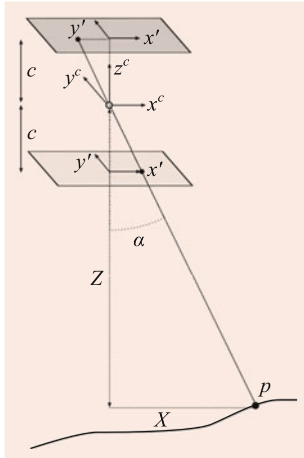


Fig. 9.26 Pinhole imaging model and its frontal counterpart: the 3-D point p is projected on the image at the intersection of the ray going through the optical center o and the image plane at a distance c. Note that $\tan(\alpha) = X/Z$

completes the right-handed system. On the image side, the distance from the optical center to the image plane along the optical axis is the camera constant c, while the distances from the intersection between the optical axis with the image plane to the image point are $-x'$ and $-y'$, respectively. By the similarity of the triangles in the upper part of Fig. 9.26 the coordinates of the image are related to that of the object by *perspective projection*.

$$x' = -c\,\frac{X}{Z}\,, \quad y' = -c\,\frac{Y}{Z}\,. \tag{9.21}$$

The negative sign in (9.21) makes the object appear upside down on the image plane. Such reversing of the scene by perspective geometry of the lens is normally compensated by the optical system, and we can eliminate this effect also mathematically by flipping the image coordinates $(-x', -y') \longmapsto (x', y')$. This is represented on the lower part of Fig. 9.26 as virtually displacing the image plane in front of the optical center, which is the so-called *frontal pinhole camera model*. We define the camera frame with x' and y'-axes identical to the *frontal* image plane, while placing its origin in the optical center. The z-axis completes the right-handed coordinate system, and its positive direction may go either "towards" or "away" from the object depending on the arbitrary choice of image coordinates. Applying the change of sign to (9.21) and combining both coordinates into a vector yields

$$\boldsymbol{x}' = \begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{c}{Z}\begin{pmatrix} X \\ Y \end{pmatrix}, \tag{9.22}$$

or equivalently in a homogeneous form

$$Z\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{9.23}$$

where $\overline{\boldsymbol{x}}' \doteq (x', y', 1)^\mathsf{T}$ and $\overline{\boldsymbol{x}}^c \doteq (X, Y, Z, 1)^\mathsf{T}$ are the homogeneous representation of image and camera coordinates, respectively. Note also that the unit of c is the same as that of x', y'.

9.3.4 Image Coordinates

In a digital camera, the measurements of features or points on the sensor are expressed in pixels. The usual convention is to situate the origin of pixel counting in the upper left-hand corner of the image and express its coordinates in terms of rows and columns. However, we need to relate the pixels to the frontal pinhole camera model. As depicted in Fig. 9.27 the optical axis intersects the sensor at the *principal point* (PP). The principal point is usually close to the physical center of the sensor denoted as the *principal point of symmetry* (PPS). Based on these different origins we define three image coordinate systems, units of which are specified in Table 9.7.

The transformation from pixel rows and columns $(\tilde{x}, \tilde{y})$ to a metric, sensor-centered image coordinate (x, y) with axis orientation as in Fig. 9.27 considers the position of PPS in pixels $(\tilde{x}_{\text{pps}}, \tilde{y}_{\text{pps}})$ and pixel size (e.g., in mm) along rows and columns $(\Delta_{\tilde{x}}, \Delta_{\tilde{y}})$

$$\begin{aligned} x &= \left[+\tilde{y} - \left(\frac{n_\text{c}}{2} - \frac{1}{2}\right)\right]\Delta_{\tilde{y}}\,, \\ y &= \left[-\tilde{x} + \left(\frac{n_\text{r}}{2} - \frac{1}{2}\right)\right]\Delta_{\tilde{x}}\,, \end{aligned} \tag{9.24}$$

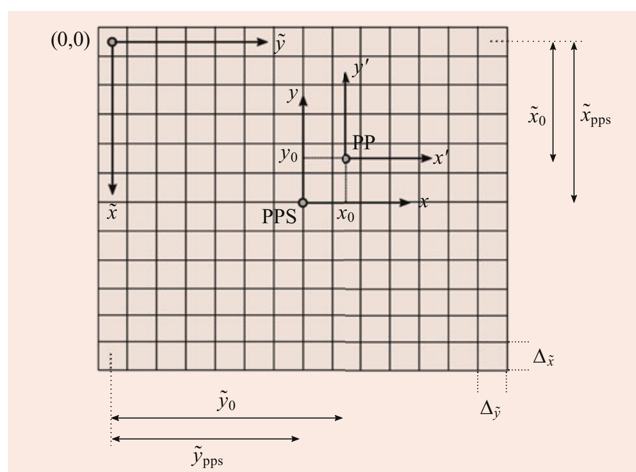


Fig. 9.27 Pixel $(\tilde{x}, \tilde{y})$, sensor-centered (x, y) and perspective-centered (x', y') image coordinates

Table 9.7 Definition of different image coordinate systems

Coordinates	Origin	Units	Usage
$(\tilde{x}, \tilde{y})$	Rows/cols counter	pixel	Computer vision
(x, y)	PPS	mm	Photogrammetry
(x', y')	PP	Unitless (= 1)	General

Table 9.8 Relations between image-coordinate systems

	$(\tilde{x}, \tilde{y})$	(x, y)	(x', y')
$(\tilde{x}, \tilde{y})$		(9.25)	(9.27)
(x, y)	(9.24)		(9.26)

where n_c and n_r are the total number of rows and columns, respectively. We express also the inverse relation, this time in a homogeneous form

$$\begin{pmatrix}\tilde{x}\\ \tilde{y}\\ 1\end{pmatrix} = \begin{pmatrix} s_x & 0 & \tilde{x}_{pps}\\ 0 & s_y & \tilde{y}_{pps}\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}-y\\ x\\ 1\end{pmatrix}, \tag{9.25}$$

with $\tilde{x}_{pps} = (n_r - 1)/2$, $\tilde{y}_{pps} = (n_c - 1)/2$ given in pixels and $s_x = 1/\Delta_{\tilde{x}}$, $s_y = 1/\Delta_{\tilde{y}}$.

To respect the perspective geometry, we define a coordinate system (x', y') with an origin placed at the principal point of autocollimation (PP). The orientation of the axis is arbitrary but in photogrammetry is usually defined as in Fig. 9.27. We also choose the unit of this coordinate system to be equal to the principal distance c, so its coordinates correspond to the tangent of angles as shown in Fig. 9.26 for x' and α. The transformation from so-called *reduced coordinates* (x', y') back to sensor-centered coordinates (x, y) in a homogeneous form is

$$\begin{pmatrix}x\\ y\\ 1\end{pmatrix} = \begin{pmatrix} c & 0 & x_0\\ 0 & c & y_0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}x'\\ y'\\ 1\end{pmatrix}, \tag{9.26}$$

with x_0, y_0 expressed in mm (sometimes displayed in μm with 1/1000 scaling factor). Analogically, the transformation from (x', y') to rows/columns $(\tilde{x}, \tilde{y})$ in pixels is

$$\begin{pmatrix}\tilde{x}\\ \tilde{y}\\ 1\end{pmatrix} = \begin{pmatrix} \tilde{c} & 0 & \tilde{x}_0\\ 0 & \tilde{c} & \tilde{y}_0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}-y'\\ x'\\ 1\end{pmatrix}, \tag{9.27}$$

with $\tilde{c}, \tilde{x}_0, \tilde{y}_0$ expressed in pixels.

Table 9.8 summarizes the transformations between the respective image-coordinate systems.

9.3.5 Imaging Formation Model

We now relate the mapping/object coordinates of point p with its coordinates on the image by means of perspective projection, while utilizing the camera frame along the way. Let us recall from Sect. 9.3.2 that the mapping coordinates of a point $\boldsymbol{x}^m = (X^m, Y^m, Z^m)^\intercal$ relative to that of a camera $\boldsymbol{x}^c$ are related by the rigid body transformation (the inverse of (9.18))

$$\overline{\boldsymbol{x}}^c = \overline{\mathbf{T}}_c^{m-1}\overline{\boldsymbol{x}}^m , \tag{9.28}$$

where the homogeneous transformation $\overline{\mathbf{T}}$ contains both the rotation and translation parameters $(\mathbf{R}, \boldsymbol{x})$.

Adopting the frontal camera model introduced in Sect. 9.3.3 for sensor-centered image coordinates (x, y) we rewrite (9.23) in vector notation

$$Z\,\overline{\boldsymbol{x}} = \begin{pmatrix} c & 0 & x_0 & 0\\ 0 & c & y_0 & 0\\ 0 & 0 & 1 & 0\end{pmatrix}\overline{\boldsymbol{x}}^c \tag{9.29}$$

Since the depth of the point p represented by Z coordinate is unknown on a single photograph, we may express it as one multiplied by an arbitrary scalar μ, i.e., $Z = \mu \cdot 1$. Decomposing the matrix in (9.29) into a product of two matrices while substituting for $\overline{\boldsymbol{x}}^c$ on the right-hand side with (9.28) we obtain the geometric model for *a basic camera*

$$\mu\begin{pmatrix}x\\ y\\ 1\end{pmatrix} = \begin{pmatrix} c & 0 & x_0\\ 0 & c & y_0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\end{pmatrix} \cdot \begin{pmatrix}\mathbf{R}_c^m & \boldsymbol{x}_c^m\\ & \\ 0 & 1\end{pmatrix}^{-1}\begin{pmatrix}X^m\\ Y^m\\ Z^m\\ 1\end{pmatrix}. \tag{9.30}$$

By defining the first two matrices on the right-hand side of the above equation as

$$\mathbf{K} \doteq \begin{pmatrix} c & 0 & x_0\\ 0 & c & y_0\\ 0 & 0 & 1\end{pmatrix}, \quad \Pi_0 \doteq \begin{pmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\end{pmatrix}, \tag{9.31}$$

we can rewrite the relation for a basic camera model (9.30) in a matrix form

$$\mu\overline{\boldsymbol{x}} = \mathbf{K}\Pi_0\overline{\boldsymbol{x}}^c = \mathbf{K}\Pi_0(\overline{\mathbf{T}}_c^m)^{-1}\overline{\boldsymbol{x}}^m = \Pi\overline{\boldsymbol{x}}^m , \tag{9.32}$$

when combining the 3×4 matrix $\mathbf{K}\Pi_0\overline{\mathbf{T}}^{-1}$ into a general *projection matrix* Π.

Now we can consider also other intrinsic parameters of a camera, for example, a basic distortion of perspective-centered image coordinates (x'_d, y'_d) with radial symmetry as

$$\begin{pmatrix}x'\\ y'\end{pmatrix} = (1 + a_1 r^2 + a_2 r^4)\begin{pmatrix}x'_d\\ y'_d\end{pmatrix}, \tag{9.33}$$

where $r^2 = x'^2_d + y'^2_d$ is the square of the distance from the center and a_1, a_2 are the *distortion coefficients*. When needed, this simple radial model can be extended by additional coefficients, as in (9.70). Combining the relation of simple image distortion (9.33) together with basic camera

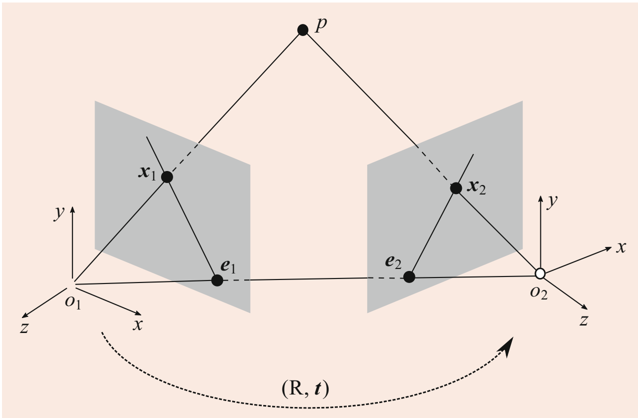


Fig. 9.28 Main steps of photogrammetry/computer-vision process in 3-D scene reconstructions based on 2-D imagery

projection model (9.32) we define the *realistic image formation model* that is applicable to many cameras employed in photogrammetry.

9.3.6 Scene from Two Views

The previously described image formation model (9.32) $\mu\overline{\boldsymbol{x}}' = \Pi\overline{\boldsymbol{x}}^m$ relates the object coordinates to image coordinates. Now we would like to perform the inverse and reconstruct 3-D object coordinates from images. As the scale (depth) μ is generally unknown due to 3-D $\longmapsto$ 2-D projection (note that μ varies per point and image), we need to employ at least two images of the same object with different camera poses that are known. Such a situation is depicted in Fig. 9.28; 3-D can be obtained by intersecting the couple of vectors pointing to the same object from two cameras. As suggested by the picture, the vector direction follows from image observation and internal camera geometry, however, both vectors need to refer to a common coordinate frame. This is the same as relating the respective camera poses to such a frame. Hence, the camera poses need to be found first. How this can be done using image observation only is described in the following.

Coplanarity Constraint

We can relate the camera pose $\overline{\mathbf{T}}_2$ to the first one $\overline{\mathbf{T}}_1$ in a relative sense $\overline{\mathbf{T}}(\mathbf{R},\boldsymbol{t}) = (\overline{\mathbf{T}}^m_{c1})^{-1}(\overline{\mathbf{T}}^m_{c2})$, so that $\overline{\boldsymbol{x}}^m_{c,2} = \overline{\mathbf{T}}\,\overline{\boldsymbol{x}}^m_{c,1}$. Expressing this in image coordinates we obtain

$$\mu_2\boldsymbol{x}_2 = \mathbf{R}\mu_1\boldsymbol{x}_1 + \boldsymbol{t}\ . \tag{9.34}$$

To eliminate the unknown depth μ we follow a couple of steps. First, we multiply both sides of the above equation from the left by a skew-symmetric matrix $[\boldsymbol{t}\times]$ containing vector $\boldsymbol{t}$. (This corresponds to cross product between two vectors.)

$$\mu_2[\boldsymbol{t}\times]\boldsymbol{x}_2 = [\boldsymbol{t}\times]\mathbf{R}\mu_1\boldsymbol{x}_1 + [\boldsymbol{t}\times]\boldsymbol{t}\ . \tag{9.35}$$

Due to orthogonality, the last term on the right-hand side is zero. Second, we multiply the last relation by $\boldsymbol{x}_2^\mathsf{T}$

$$\mu_2\boldsymbol{x}_2^\mathsf{T}[\boldsymbol{t}\times]\boldsymbol{x}_2 = \boldsymbol{x}_2^\mathsf{T}[\boldsymbol{t}\times]\mathbf{R}\mu_1\boldsymbol{x}_1\ . \tag{9.36}$$

Since $[\boldsymbol{t}\times]\boldsymbol{x}_2$ is perpendicular to $\boldsymbol{x}_2$, and the inner product of the two perpendicular vectors is zero, the left-hand side $\boldsymbol{x}_2^\mathsf{T}[\boldsymbol{t}\times]\boldsymbol{x}_2 = 0$. Also, as $\mu_1 \neq 0$, we can write

$$\boldsymbol{x}_2^\mathsf{T}[\boldsymbol{t}\times]\mathbf{R}\,\boldsymbol{x}_1 = \boldsymbol{x}_2^\mathsf{T}\mathbf{E}\,\boldsymbol{x}_1 = 0\ , \tag{9.37}$$

where $\mathbf{E} \doteq [\boldsymbol{t}\times]\mathbf{R}$ is called the *essential matrix*. The above relation is called an *epipolar constraint* as it conditions the three vectors $\boldsymbol{x}_2$, $\boldsymbol{t}$, and $\mathbf{R}\,\boldsymbol{x}_1$ to lie on a common plane, denoted as *epipolar plane*. Fig. 9.28 depicts also the two *epipoles* $\boldsymbol{e}_1, \boldsymbol{e}_2$ resulting from the intersection between a line $\mathbf{o}_1$–$\mathbf{o}_2$ and respective image planes. Connections $\boldsymbol{e}_1$–$\boldsymbol{x}_1$ and $\boldsymbol{e}_2$–$\boldsymbol{x}_2$ are called *epipolar lines*.

Essential Matrix Determination

To reconstruct $\mathbf{E}$ using only image observations, we briefly present the basic algorithm of [43] known also as the eight-point algorithm.

First, we stack the 3×3 entries of $\mathbf{E}$ into a vector by columns i.e., $E^s \doteq (e_{11}, e_{21}, e_{31}, e_{12}, \ldots, e_{33})^\mathsf{T}$. Our goal is to determine this vector and obtain $\mathbf{E}$ by its "unstacking".

Second, we make use of the *Kronecker product* of two vectors,

$$\boldsymbol{a} \doteq \boldsymbol{x}_1 \otimes \boldsymbol{x}_2 = (x_1\boldsymbol{x}_2, x_2\boldsymbol{x}_2, x_3\boldsymbol{x}_2)^\mathsf{T}\ , \tag{9.38}$$

to express the epipolar constraint per one correspondence (point) as

$$\boldsymbol{a}^\mathsf{T} E^s = 0\ . \tag{9.39}$$

Having a set of corresponding image points $(\boldsymbol{x}_1^i, \boldsymbol{x}_2^i)$, $i = 1, 2, \ldots, n$ we create n vectors $(\boldsymbol{a}^i)^\mathsf{T}$ and put them into a matrix $\chi \doteq (\boldsymbol{a}^1; \boldsymbol{a}^2; \ldots; \boldsymbol{a}^n)$. With that we express the epipolar conditions for all correspondences in a system of linear equations

$$\chi E^s = 0\ . \tag{9.40}$$

The solution of this equation is unique if the rank of the matrix χ is exactly 8 (a global scale factor cannot be determined). For this reason, we need $n \geq 8$ points. Note that expression (9.40) holds only in the absence of noise. However, in reality, we have to deal with noise and we are likely to have more correspondences. In the eight-point algorithm, the choice is made to minimize the least-squares error function of *misclosures* $|\boldsymbol{\chi} E^s|^2 \neq 0$, which is achieved by choosing E^s to be an eigenvector of $(\chi^\top \chi)$ that corresponds to its smallest singular value (eigenvalue) λ. Practically, this can be found by performing a singular value decomposition of $\boldsymbol{\chi} = \mathbf{U}_\chi \boldsymbol{\Sigma}_\chi \mathbf{V}_\chi^\top$; i.e., factoring χ into a product of diagonal matrix $\boldsymbol{\Sigma}_\chi$ containing the eigenvalues and orthogonal matrices $(\mathbf{U}^\top\mathbf{U} = \mathbf{I})$ $\mathbf{U}_\chi$ and $\mathbf{V}_\chi$; and defining E^s to be the column of $\mathbf{V}_\chi$ associated with the smallest singular value. Then we reshape the nine elements of E^s into 3×3 matrix $\mathbf{E}$.

While the reconstructed $\mathbf{E}$ minimizes the norm $|\boldsymbol{\chi} E^s|^2$ in the least-squares sense, it is not guaranteed – due to the observation of unmodeled errors – that its structure belongs to the space of essential matrices. This space is characterized by $\mathbf{E} = \mathbf{U}\,\mathrm{diag}\{\sigma, \sigma, 0\}\mathbf{V}^\top$, where $\sigma = |\boldsymbol{t}|$. A common approach is, therefore, to reproject the estimated $\mathbf{E}$ to such a space. This is achieved by carrying the singular value decomposition of $\mathbf{E}$

$$\mathbf{E} = \mathbf{U}\,\mathrm{diag}\{\lambda_1, \lambda_2, \lambda_3\}\mathbf{V}^\top\,, \tag{9.41}$$

with $\lambda_1 > \lambda_2 > \lambda_3 \neq 0$ and then setting the smallest eigenvalue to zero and other two as $0.5(\lambda_1 + \lambda_2)$. Alternatively, as the global scale cannot be recovered by image observations only, it may well be chosen as unity, which corresponds to *normalized essential space* where the two largest eigenvalues are set to 1.

Pose Reconstruction

Let us define a rotation matrix

$$\mathbf{R}_z(\pm\pi/2) = \begin{pmatrix} 0 & \mp 1 & 0 \\ \pm 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{9.42}$$

where the meaning of $\pm$ and $\mp$ signs will be explained later. Considering that (as for any rotation matrix) $\mathbf{R}_z\mathbf{R}_z^\top = \mathbf{I}$ together with the elements of singular value decomposition of $\mathbf{E}$, we can verify the correctness of the following relation

$$\begin{aligned}\mathbf{E} = [\boldsymbol{t}\times]\mathbf{R} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top &= \mathbf{U}\mathbf{R}_z\boldsymbol{\Sigma}\mathbf{R}_z^\top\mathbf{V}^\top \\ &= \underbrace{\mathbf{U}\mathbf{R}_z\boldsymbol{\Sigma}\mathbf{U}^\top}_{[\boldsymbol{t}\times]}\underbrace{\mathbf{U}\mathbf{R}_z^\top\mathbf{V}^\top}_{\mathbf{R}}\,,\end{aligned} \tag{9.43}$$

since $\mathbf{U}^\top\mathbf{U} = \mathbf{I}$ due to the orthonormality of $\mathbf{U}$ after reprojection of $\mathbf{E}$. Then the relative rotation follows directly from (9.43) as the product of three rotation matrices on the right, i.e.,

$$\mathbf{R} = \mathbf{U}\mathbf{R}_z^\top(\pm\pi/2)\mathbf{V}^\top\,, \tag{9.44}$$

and the relative translation (up to a scale) as

$$[\boldsymbol{t}\times] = \mathbf{U}\mathbf{R}_z^\top(\pm\pi/2)\boldsymbol{\Sigma}\mathbf{U}^\top\,, \tag{9.45}$$

where it can be proven that $\mathbf{U}\mathbf{R}_z^\top(\pm\pi/2)\boldsymbol{\Sigma}\mathbf{U}^\top$ is of a skew-symmetric form. The $\pm$ sign in $\mathbf{R}_z(\pm\pi/2)$ reflects the fact that each essential matrix gives two possible solutions and its reconstructed sign is arbitrary. Hence, we could possibly obtain up to four solutions of the relative pose $(\mathbf{R}, \boldsymbol{t})$ from $\pm\mathbf{E}$. Three of these correspond to situations where the scene captured by one camera (first or second) or both cameras is situated behind the lens, which is physically not possible. These solutions can be eliminated by imposing the positive depth constraint. We should also mention that despite its simplicity, the eight-point algorithm is not without potential numerical weaknesses that may become apparent in a particular geometry and observation noise. However, as demonstrated by [44], these can be avoided by data preprocessing (translation and scaling).

Structure Reconstruction

With eight or more correspondences as an input, the previously described algorithm determined the relative rotation and translation between the two cameras, the latter up to a global scale (ξ). Setting the norm of the translation vector to unity is equivalent to choosing $\xi = 1$. The relative pose can then be used to retrieve the position of the other correspondences on the images in 3-D.

Considering again the relation (9.34), which relates camera poses to n image correspondences

$$\mu_2^i \boldsymbol{x}_2^i = \mu_1^i \mathbf{R}\boldsymbol{x}_1^i + \xi\boldsymbol{t}\,, \quad i = 1, 2, \ldots, n\,. \tag{9.46}$$

Since $(\mathbf{R}, \boldsymbol{t})$ are known, this relation is linear and can, therefore, easily be solved once the unknown depth μ_1, μ_2 with respect to the first and second camera frames are determined. One of them is, however, redundant, as it is function of $(\mathbf{R}, \boldsymbol{t})$, as well as the arbitrary choice of the global scale ξ. Hence, we can eliminate, for instance, μ_2 by multiplying the above equation by the orthogonal operator $[\boldsymbol{x}_2\times]$ to obtain

$$\mu_1^i[\boldsymbol{x}_2^i\times]\mathbf{R}\boldsymbol{x}_1^i + \xi[\boldsymbol{x}_2^i\times]\boldsymbol{t} = 0\,. \tag{9.47}$$

An equivalent form that regroups the unknowns in a common vector is

$$\left([\boldsymbol{x}_2^i\times]\mathbf{R}\boldsymbol{x}_1^i, \quad [\boldsymbol{x}_2^i\times]\boldsymbol{t}\right)\begin{pmatrix}\mu_1^i \\ \xi\end{pmatrix} \doteq \mathbf{M}^i\boldsymbol{\mu}^i = 0\,. \tag{9.48}$$

To obtain a unique solution the matrix $\mathbf{M}$ needs to be of rank 1 or $[\boldsymbol{x}_2^i\times]\boldsymbol{t} \neq 0$. Notice that this is not the case when the point p lies on the line connecting two optical centers.

Regrouping all n correspondences into one equation while noticing that ξ is common to all of them we obtain $\overline{\boldsymbol{\mu}} =$

9

$(\mu_1^1, \mu_1^2, \ldots, \mu_1^n, \xi)^\mathsf{T}$ and a matrix $\mathbf{M}$ defined as

$$\mathbf{M} \doteq \begin{pmatrix} [\boldsymbol{x}_2^1\times]\mathbf{R}\boldsymbol{x}_1^1 & 0 & 0 & [\boldsymbol{x}_2^1\times]\boldsymbol{t} \\ 0 & [\boldsymbol{x}_2^2\times]\mathbf{R}\boldsymbol{x}_1^2 & 0 & [\boldsymbol{x}_2^2\times]\boldsymbol{t} \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & [\boldsymbol{x}_2^n\times]\mathbf{R}\boldsymbol{x}_1^n & [\boldsymbol{x}_2^n\times]\boldsymbol{t} \end{pmatrix} . \tag{9.49}$$

The solution to the equation

$$\mathbf{M}\overline{\boldsymbol{\mu}} = 0 \tag{9.50}$$

determines all the unknowns in the vector $\overline{\boldsymbol{\mu}}$ up to the last one corresponding to the one global scale ξ. Similarly to the approach of essential matrix determination, the minimization of the square of misclosures (9.50) can be found as the eigenvector of $\mathbf{M}^\mathsf{T}\mathbf{M}$ that corresponds to the smallest eigenvalue.

Global Scale

The global scale can be determined only by some exterior knowledge either on the camera motion, as discussed in Sect. 9.4.4, or on the object coordinates of some observed points. For instance, if the restitution of structure is required in a mapping frame, we need to know the coordinates of at least three points in both frames to apply the seven-parameter similarity transformation. Finally, the problem of reconstruction can be formulated as an unconstrained optimization problem, where the minimization is searched with respect to all unknowns $\boldsymbol{x}_1^i, \mathbf{R}, \boldsymbol{t}, \bar{\mu}$. This is known in the literature as bundle adjustment, and its method will be further detailed in the next section, as well as in Sect. 9.4.3. The presented form, however, allows us to develop the needed approximations for its effective solution, which are based exclusively on image observations.

9.3.7 Scene from Multiple Views

Multiple-View Matrix

We now consider the existence of more than two views of the same object, which is rather a standard case. Without loss of generality, let us take the frame of the first camera as a reference frame for 3-D reconstruction. With m views/images at our disposition, from (9.32), we obtain the following projection matrices

$$\Pi_1 = (\mathbf{I}, \mathbf{0}), \Pi_2 = (\mathbf{R}_2, \boldsymbol{t}_2), \ldots, \Pi_m = (\mathbf{R}_m, \boldsymbol{t}_m) , \tag{9.51}$$

Considering at the moment only one point p and applying a similar development as for relations (9.46)–(9.48) we can derive the *multiple-view matrix* $\mathbf{M}_p$. We do so by inserting into two columns of $\mathbf{M}_p$ a coplanarity constraint (9.34) of view i between the first and i-th camera reference frame. Up to its depth this constraint is $\mu_1[\boldsymbol{x}_i\times]\mathbf{R}_i\boldsymbol{x}_1 + [\boldsymbol{x}_i\times]\boldsymbol{t}_i = 0$. In matrix form this is

$$\mathbf{M}_p \begin{pmatrix} \mu_1 \\ 1 \end{pmatrix} = 0 , \tag{9.52}$$

with $\mathbf{M}_p$ defined as

$$\mathbf{M}_p \doteq \begin{pmatrix} [\boldsymbol{x}_2\times]\mathbf{R}_2\boldsymbol{x}_1 & [\boldsymbol{x}_2\times]\boldsymbol{t}_2 \\ [\boldsymbol{x}_3\times]\mathbf{R}_3\boldsymbol{x}_1 & [\boldsymbol{x}_3\times]\boldsymbol{t}_3 \\ \vdots & \vdots \\ [\boldsymbol{x}_m\times]\mathbf{R}_m\boldsymbol{x}_1 & [\boldsymbol{x}_m\times]\boldsymbol{t}_m \end{pmatrix} . \tag{9.53}$$

This matrix thus associates m views of point p by involving both the image $\boldsymbol{x}_1$ and the coimages $[\boldsymbol{x}_2\times], [\boldsymbol{x}_3\times], \ldots, [\boldsymbol{x}_m\times]$. In other words, it encodes all constraints that exist among the m views of a point. It has rank 1, as long as the pair of vectors $[\boldsymbol{x}_i\times]\boldsymbol{t}_i$, $[\boldsymbol{x}_i\times]\mathbf{R}_i\boldsymbol{x}_i$ is linearly dependent for each $i = 1, 2, \ldots, m$, which is equivalent to the *bilinear epipolar constraints*

$$\boldsymbol{x}_i^\mathsf{T}[\boldsymbol{t}_i\times]\mathbf{R}_i\boldsymbol{x}_1 = 0 . \tag{9.54}$$

In such a situation, the projection of p on the image is unique, which is not the case for $\texttt{rank}(\mathbf{M}_p) = 2$ or $\texttt{rank}(\mathbf{M}_p) = 0$. Rank testing can potentially be used for filtering out the wrongly established correspondences in feature matching.

Trilinear Constraint

In some situations, it may be useful to formulate one condition involving directly three views. Let us consider one point that is viewed by three cameras $1, i, j$. For this situation, we can write two separate coplanarity constraints, the second being transposed

$$\begin{aligned} \mu_1[\boldsymbol{x}_i\times]\mathbf{R}_i\boldsymbol{x}_1 &= -[\boldsymbol{x}_i\times]\boldsymbol{t}_i , \\ \boldsymbol{x}_1^\mathsf{T}\mathbf{R}_j^\mathsf{T}[\boldsymbol{x}_j\times]^\mathsf{T}\mu_1 &= -\boldsymbol{t}_j^\mathsf{T}[\boldsymbol{x}_j\times]^\mathsf{T} . \end{aligned} \tag{9.55}$$

Multiplying across the left and right-hand sides of both equations (9.55) and making them equal

$$-[\boldsymbol{x}_i\times]\mathbf{R}_i\boldsymbol{x}_1\boldsymbol{t}_j^\mathsf{T}[\boldsymbol{x}_j\times]^\mathsf{T} = -[\boldsymbol{x}_i\times]\boldsymbol{t}_i\boldsymbol{x}_1^\mathsf{T}\mathbf{R}_j^\mathsf{T}[\boldsymbol{x}_j\times]^\mathsf{T} , \tag{9.56}$$

then rearranging the terms to one side, we obtain the *trilinear constraint*

$$[\boldsymbol{x}_i\times]\left(\boldsymbol{t}_i\boldsymbol{x}_1^\mathsf{T}\mathbf{R}_j^\mathsf{T} - \mathbf{R}_i\boldsymbol{x}_1\boldsymbol{t}_j^\mathsf{T}\right)[\boldsymbol{x}_j\times] = 0 . \tag{9.57}$$

The trilinear constraint implies a bilinear constraint (9.54), except for a special case in which $[\boldsymbol{x}_j\times]\boldsymbol{t}_j = [\boldsymbol{x}_j\times]\mathbf{R}_j\boldsymbol{x}_i = 0$ for some view j. In this rare situation, the point p lies on the line connecting the optical centers $\mathbf{o}_1, \mathbf{o}_j$. The application of the trilinear constraint may, therefore, be of a certain advantage for some special cases, such as that when three image vectors of the same point are coplanar. (This may be the case, for instance, in a car-based mapping system when views from the same forward-looking camera are combined between successive times, i.e., involving displacement only along the depth of field.) When they still satisfy the trilinear constraint, 3-D coordinates of this point can be reconstructed. It should also be mentioned that any other algebraic constraint among m images can be reduced to those involving either two or three at a time (i.e., application of either bilinear or trilinear constraints).

Processing Strategies

The processing strategies for handling multiple views vary in terms of image geometry, scene texture (goodness of feature detection, matching, and filtering), camera calibration, data noise, and experience. We therefore present only the main concepts while leaving out the details of their combination into to a particular implementation. These are schematically depicted in Fig. 9.29.

In principle, any multiple view can be broken down into a sequence of two-view scenarios between first and last camera poses. This situation is highlighted in the upper part of Fig. 9.29 and is often used in practice when the overlap between images is small; the texture allows finding only few correspondences, or there is a large uncertainty in the camera model. To mitigate the accumulation of random influences in the sequence, the "two-view step" is followed by a global optimization involving all views.

The second approach is to use eight-point algorithm only once for some initial pairs of view and perform global optimization on the rest, as shown in the bottom part of Fig. 9.29. This method is more suitable for good image geometry, large overlap, and (pre)calibrated cameras. It involves constructing a relation containing the multiple-view matrix $\mathbf{P}_i$, similar to that of (9.53) but involving m images $\boldsymbol{x}_1^j, \boldsymbol{x}_2^j, \ldots, \boldsymbol{x}_1^m$ of n points p^j, $j = 1, 2, \ldots n$ from which we would like to estimate the unknown projections $\Pi_i(\mathbf{R}_i^s, \boldsymbol{t}_i)^\mathsf{T}$, $i = 2, 3, \ldots, m$

$$\mathbf{P}_i \begin{pmatrix} \mathbf{R}_i^s \\ \boldsymbol{t}_i \end{pmatrix} \begin{pmatrix} [\boldsymbol{x}_1^1 \times]^\mathsf{T} \otimes [\boldsymbol{x}_i^1 \times] & \lambda^1 [\boldsymbol{x}_i^1 \times] \\ [\boldsymbol{x}_1^2 \times]^\mathsf{T} \otimes [\boldsymbol{x}_i^2 \times] & \lambda^2 [\boldsymbol{x}_i^2 \times] \\ \vdots & \vdots \\ [\boldsymbol{x}_1^n \times]^\mathsf{T} \otimes [\boldsymbol{x}_i^n \times] & \lambda^n [\boldsymbol{x}_i^n \times] \end{pmatrix} \begin{pmatrix} \mathbf{R}_i^s \\ \boldsymbol{t}_i \end{pmatrix} = 0\,, \tag{9.58}$$

where $\otimes$ is the *Kronecker product* between matrices (similarly to that of (9.38) but with the elements of the left matrix stacked into a vector), and λ^j is the inverse of unknown depth μ^i. The matrix $\mathbf{P}_i$ is of size $3n \times 12$ and is of rank 11 if more than $n \geq 6$ points are provided. Then the projection matrix $\Pi_i = (\mathbf{R}_i, \boldsymbol{t}_i)$ can be solved for up to the scale factor. When $\mathbf{P}_i$ is of rank higher than 11, the solution that minimizes the square of misclosures is obtained as an eigenvector of $\mathbf{P}_i$ associated with the largest eigenvalue. As the estimate of Π_i is affected by random errors, the estimated matrix $\mathbf{R}_i$ needs to be reprojected to the rotational space $SO(3)$ and the vector $\boldsymbol{t}_i$ rescaled. Assuming the pose for the second view is found by the eight-point algorithm, the scalars λ^j can be determined from the first row of (9.58) involving $\lambda^j [\boldsymbol{x}_2^j \times] \boldsymbol{t}_2 = -[\boldsymbol{x}_2^j \times] \mathbf{R}_2 \boldsymbol{x}_1^j$. These initial values of λ^j can then be used for the recovery of Π_i, $i = 3, 4, \ldots, m$. Since $\boldsymbol{t}_2$ is recovered from the eight-point algorithm up to a scale factor, the other views are recovered from that up to a global single scale.

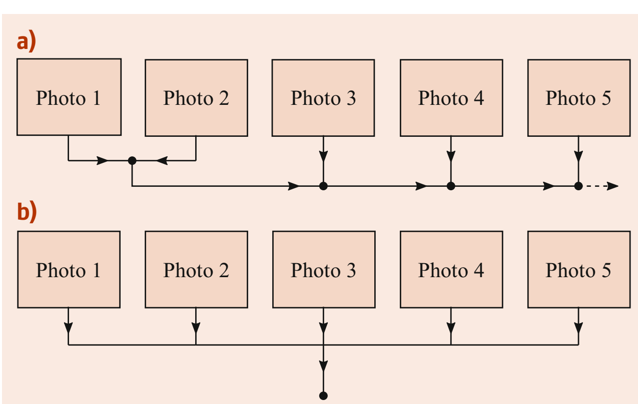


Fig. 9.29 Image reconstruction strategies: incremental (**a**) versus global (**b**)

Optimization

The equivalent and perhaps even simpler formulation of the global optimization, i.e., the concurrent determination of object coordinates, camera parameters, and pose (i.e., structure and motion) is presented under a name of *bundle adjustment*, which extension also accommodates other inputs, is presented in Sect. 9.4 on sensor fusion. Bundle adjustment received its name after application of the ray-tracing collinearity condition (9.32) $\mu \overline{\boldsymbol{x}'} = \Pi \overline{\boldsymbol{x}}^m$ on a "bundle of rays" connecting object points with its projection on the image in combination with a particular sensor model (9.33). Nevertheless, this optimization approach requires linearization and an existence of approximate values of parameters. With an exclusive use of image observations, the approximate value of parameters can be obtained by the methods described in this and previous sections.

9.3.8 Feature Matching

The term image matching stands for the mostly automatic reference between regions or pixels of two or more images that represent the same feature or point in the object space. Automatic aerial triangulation (AAT) requires the availability of suitable image matching tools as a key component. These tools should enable fully automatic tie-point measurement by providing homologous features with suitable accuracy and reliability. For this purpose, feature-based matching approaches are frequently used. First, primitives suitable for image matching are extracted, while in a second step, their correspondences are determined by some similarity and consistency measures. These two steps of feature-based matching techniques result in a categorization into feature detectors and feature descriptors. Detectors search for image points or regions that are geometrically stable under different transformations and have a high information content. The results are generally called interest points, corners, or invariant regions. Descriptors instead analyze the image to provide a 2-D vector of image information at those areas defined by the respective interest point. The subsequent matching process then exploits this information for similarity measurement in order to evaluate potential point correspondences. To remove outliers remaining after this matching, geometric constraints such as epipolar geometry are applied by robust estimators in a final step.

Feature extraction and matching are strongly related; however, these two steps are discussed separately in the next sections. This separation also results from the high accuracy demands within automatic aerial triangulation, which is usually fulfilled by hybrid matching approaches. In this context, tie-point positions as provided by feature-based methods are refined in a subsequent step using intensity-based correlation strategies.

Image Matching Primitives

To detect primitives suitable for image matching, so-called interest operators were first developed in the 1970s. Since then, a wide variety of algorithms have evolved in computer vision, pattern recognition, and photogrammetry. Comprehensive overviews on feature extraction are given, e.g., in [45, 46]. In the context of image matching, feature extraction aims to identify primitives, which are invariant against radiometric and geometric distortions, robust against image noise, and distinguishable from other points [47]. This task is especially complex for close-range applications in which one frequently has to cope with convergent images with different look angles at varying scale. However, the situation is easier for aerial triangulation. In this context, similar viewpoints and relatively short time intervals during image collection avoid problems due to perspective distortions and large changes in illumination. Furthermore, matching can be simplified using a-priori information on the respective image geometry, which is usually available from camera calibration, the standardized flight geometry of airborne image blocks, or measured GNSS trajectories. Within commercially available AAT software, the Förstner operator [48] has been widely used. This operator was developed for fast detection and precise location of distinct points including corners and centers of circular image features.

Feature detectors such as the Förstner and Harris operators were mainly integrated for applications in airborne photogrammetry [49]. Meanwhile, the scale-invariant feature transform (SIFT) keypoint detector [50] has become the quasi standard for point extraction and matching. Scale invariant means that a feature in object space that appears with a large scale in one image and with a small scale in another can still be detected as the same by the SIFT-operator. It is scale invariant since feature points are detected in the so-called scale space by searching for maxima in an image pyramid as defined by a stack of the difference of Gaussians (DoG) [50]. Thus, it has become especially popular in close-range applications, where matching is frequently aggravated due to the appearance of larger perspective distortions.

Feature Matching Strategies

Feature detection is followed by a suitable matching step to provide the required point correspondences for the aspired AAT. This matching is based on information representing the local image patch in the vicinity of the respective feature point. Attributes are usually derived from the gray or gradient values in the feature's neighborhood. As an example, the feature description for the SIFT operator is generated from the histogram of the gradient vectors in the local neighborhood of the keypoint location [50]. This approach transforms the image data into a scale and rotation-invariant representation. A pair of keypoints within two overlapping images is then regarded as corresponding if the Euclidian distance between their respective descriptors is less than a given threshold, and the distance to the second nearest descriptor is greater than a second given threshold. An overview on the use of local descriptors is given in [51].

If feature matching is required during the evaluation of aerial imagery, the homogeneity conditions during image collection usually allow for the use of gray values in the local vicinity of a feature point. Thus, the similarities of potentially corresponding image patches can be measured by normalized cross correlation (NCC)

$$\rho(r,c) = \frac{\begin{pmatrix} \sum_{i=1}^{m}\sum_{j=1}^{n}[g_\mathrm{L}(i,j) - \bar{g}_\mathrm{L}] \\ \cdot [g_\mathrm{R}(r+i,c+j) - \bar{g}_\mathrm{R}] \end{pmatrix}}{\sqrt{\begin{matrix} \sum_{i=1}^{m}\sum_{j=1}^{n}[g_\mathrm{L}(i,j) - \bar{g}_\mathrm{L}]^2 \\ \cdot \sum_{i=1}^{m}\sum_{j=1}^{n}[g_\mathrm{R}(r+i,r+j) - \bar{g}_\mathrm{R}]^2 \end{matrix}}}\,, \tag{9.59}$$

where ρ ist the normalized cross correlation, r and c indicate row and column, m and n the shift of rows and columns between both images, g_L, g_R the gray values of a pixel in left and right images, and $\bar{g}_L, \bar{g}_R$ the average gray values of the search window in the left and right images.

This provides values normalized in the interval with highest similarity for a coefficient close to 1. Usually, such similarity measurements are not robust enough to avoid mismatches. Hence, an additional step to reject potential outliers is required. For this purpose, geometric constraints as provided from the epipolar geometry of the respective image pair are frequently used. As an example, algorithms based on random sample consensus (RANSAC) [52] robustly estimate the relative orientation between image pairs.

This provides a suitable transformation for the corresponding image points and, therefore, allows elimination of potential mismatches while providing consistent point correspondences. The algorithm can be summarized as follows:

1. A random sample of five corresponding image points is taken from the list of matched points of the two images.
2. From these five correspondences, the relative orientation of the image pair is computed using the algorithm described in [53], or, alternatively, with (9.43) when setting the translation vector to unity.
3. This relative orientation defines for each image point $j = 1, 2, \ldots, n$ in the left image the corresponding epipolar line in the right image (9.37), $\boldsymbol{x}_2^j \mathbf{E} \boldsymbol{x}_1^j = 0$. The difference between this epipolar line and the corresponding point in the right image defines an error for this potential match with respect to the calculated relative orientation. If a matched point pair has a small epipolar error, it fits well with the estimated relative orientation. In that case, this potential correspondence is considered as a hypothetical inlier, otherwise it is an outlier as schematically depicted in Fig. 9.30.
4. If sufficiently many point pairs are classified as inliers, the estimated relative orientation is reasonably good, and the algorithm can be terminated. All inliers are preserved, while the outliers are eliminated from the final list of correspondences.
5. Otherwise, the RANSAC algorithm continues with step 1 with another random sample of five corresponding image points.

These correspondences can be directly used as tie-points during optimization with other data. However, the accuracy of the applied feature-based matching is usually increased to subpixel level by subsequent area-based matching. This can, e.g., be realized using NCC as defined by (9.59). For subpixel measurement, the center of the correlation masks g_L and g_R are again defined by the coordinates of the left and right feature points. The NCC is then computed in a local 3×3 neighborhood of the potential match. Of course, the best similarity position defined by the maximal NCC coefficient will correspond to the center point of this matrix, i.e., the coordinates of the corresponding right feature point. However, the correlation coefficients in the local neighborhood of this best match position can be used for subpixel refinement by interpolation through the second-order polynomial. The cross sections in row and column directions are parabolas.

$$f(r, c) = a_0 + a_1 r + a_2 c + a_3 r^2 + a_4 rc + a_5 c^2 . \quad (9.60)$$

$$\begin{pmatrix} \rho_0(-1,-1) & \rho_1(-1,0) & \rho_2(-1,1) \\ \rho_3(0,-1) & \rho_4(0,0) & \rho_5(0,1) \\ \rho_6(1,-1) & \rho_7(1,0) & \rho_8(1,1) \end{pmatrix} , \quad (9.61)$$

which represents the NCC coefficients for a $\times$ local neighborhood centered at position $(0, 0)$ of the maximum value p_4. The NCC coefficients $\boldsymbol{l} = (\rho_0, \rho_1, \ldots, \rho_8)^\top$ computed by (9.59) for the different positions (r_i, c_i) are then used as observations within the Gauss–Markov model $\mathbf{A}\boldsymbol{x} - \boldsymbol{l} = \boldsymbol{v}$. The parameters of the polynomial (9.60) can then be estimated with

$$\mathbf{A} = \begin{pmatrix} 1 & r_0 & c_0 & r_0^2 & r_0 c_0 & c_0^2 \\ 1 & r_1 & c_1 & r_1^2 & r_1 c_1 & c_1^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & r_8 & c_8 & r_8^2 & r_8 c_8 & c_8^2 \end{pmatrix} . \quad (9.62)$$

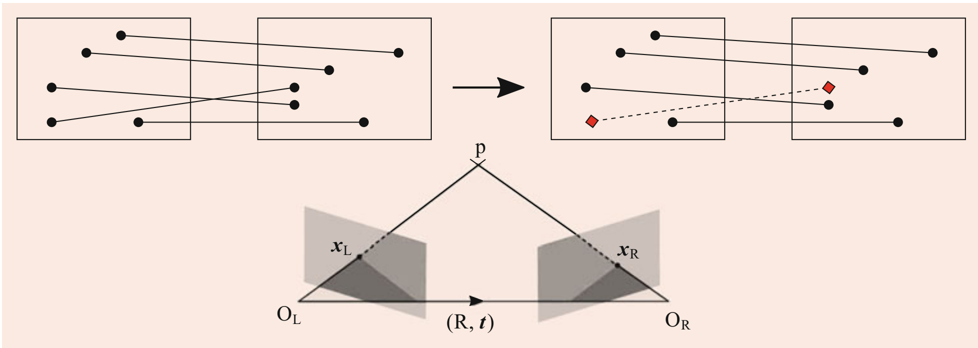


Fig. 9.30 Filtering of wrongly assigned correspondences on a randomly selected subset of points through the epipolar constraint

If the available values for (r_i, c_i) are introduced, the Gauss–Markov model results in the equation

$$\begin{pmatrix} 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 1 & 1 & -1 & 1 \\ 1 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} - \begin{pmatrix} \rho_0 \\ \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ \rho_5 \\ \rho_6 \\ \rho_7 \\ \rho_8 \end{pmatrix} = V \,. \quad (9.63)$$

The standard solution

$$\boldsymbol{x} = (\mathbf{A}^{\mathsf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathsf{T}}\boldsymbol{l} = [\mathbf{A}(\mathbf{A}^{\mathsf{T}}\mathbf{A})^{-1}]^{\mathsf{T}}\boldsymbol{l} = \boldsymbol{t}'\boldsymbol{l}$$

then provides the five parameters of the polynomial $\boldsymbol{x} = (a_0, a_1, \ldots, a_4)^{\mathsf{T}}$. The partial derivatives then define the extremum of this polynomial in row and column directions by

$$\frac{\partial f}{\partial r} = a_1 + 2a_3\Delta r + a_4\Delta c \cong 0 \,, \quad (9.64)$$

$$\frac{\partial f}{\partial c} = a_2 + a_4\Delta r + a_5\Delta c \cong 0 \,. \quad (9.65)$$

This finally gives the subpixel refinement $(\Delta r, \Delta c)$ for the initial center position of the best match as

$$\Delta r = \frac{a_2 a_4 - 2a_1 a_5}{4a_3 a_5 - a_4^2} \,, \quad (9.66)$$

$$\Delta c = \frac{a_1 a_4 - 2a_2 a_3}{4a_3 a_5 - a_4^2} \,. \quad (9.67)$$

An example input for the computation of subpixel refinement can be found in Fig. 9.31, which depicts a correlation mask in the left-hand image and a search mask in the right-hand image, both centered at their corresponding feature point. The NCC coefficients ρ_i computed from using the correlation mask within the 3×3 neighborhood of the right-hand feature point give the matrix

$$\begin{pmatrix} 0.9300 & 0.9088 & 0.8622 \\ 0.9862 & 0.9922 & 0.9664 \\ 0.9281 & 0.9646 & 0.9696 \end{pmatrix} . \quad (9.68)$$

If these values are used as input for (9.63) and (9.66), this gives a shift of $\Delta r = -0.0409$ pixels and $\Delta c = 0.2366$ pixels.

The use of local sample points to determine the interpolated location of the maximum has also been used in the context of SIFT keypoint extraction [50]. In this application, pixel coordinates x, y and scale s define a 3-D scale space function $D(x, y, s)$. Thus, subpixel and subscale coordinates $x = (x; y; s)$ of a feature are found by interpolation with a 3-D quadratic function that has the shape of a parabola in each of the three dimensions. This provides a substantial improvement to matching and stability.

As an alternative to NCC-based subpixel measurement of tie-point coordinates, least-squares image matching [48] can be applied. This approach estimates the geometric and radiometric transformations between corresponding patches g_L and g_R from the left-hand and right-hand images, respectively, using the observation equation

$$g_L(r,c) + v = h_1 + h_2 g_R[(a_0 + a_1 r + a_2 c), (b_0 + b_1 r + b_2 c)] \,. \quad (9.69)$$

This approach models geometric differences between image patches by a simple affine transformation with parameters a_0, a_1, a_2, b_0, b_1, and b_2, while radiometric differences caused, e.g., by different sun lighting are represented by offset and gain, h_1 and h_2. The transformation parameters are then estimated through iterate least-squares adjustment.

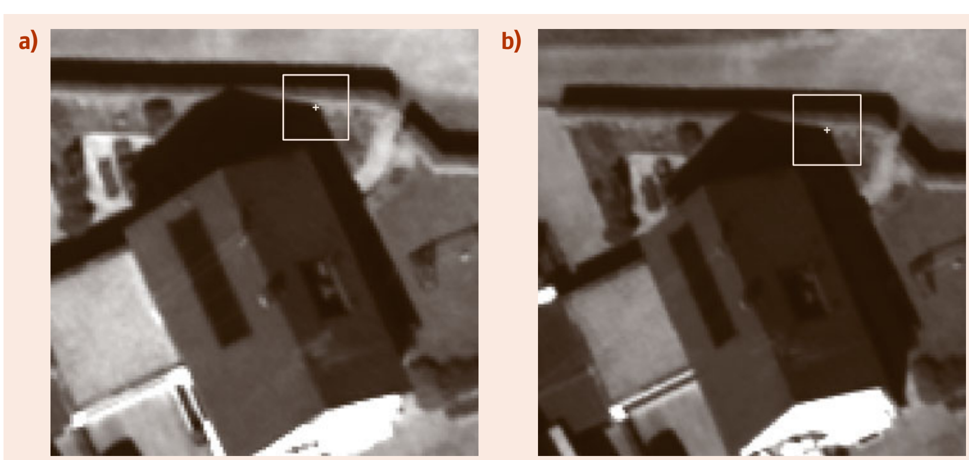


Fig. 9.31 Feature point with correlation mask (**a**) and search mask (**b**)

Table 9.9 Participants of the EuroSDR-Project Benchmark on Image Matching, 2014

Name of Software	Manufacturer	Location
SocetSet 5.6 (NGATE)	BAE Systems	Newcastle upon Tyne, UK
UltraMap V3.1	Microsoft, Vexcel	Graz, Austria
Match-T DSM 5.5	Trimble/inpho	Stuttgart, Germany
ImageStation ISAE-Ext	GEOSYSTEMS GmbH	Munich, Germany
Pixel Factory	Astrium GEO-Information Services	Paris, France
RMA DSM Tool	Royal Military Academy (RMA)	Brussels, Belgium
Remote-sensing software package	Joanneum Research	Graz, Austria
MicMac	IGN France	Paris, France
SURE	IfP, University of Stuttgart	Stuttgart, Germany
FPGA implementation of his SGM	German Aerospace Center (DLR)	Oberpfaffenhofen, Germany
XProSGM	Leica	Heerbrugg, Switzerland

Equation 9.69 equals the Gauss–Markov model, which minimizes the squared sum of errors of all the observations $v^{\mathsf{T}}v \rightarrow \min$. Gray-value differences between the corresponding image patches are used for a typical window size of 15×15 pixel, resulting in a matching precision of 0.1–0.01 pixels.

Dense Matching

Stereo matching aiming at the automatic generation of elevation data from aerial images was already introduced more than two decades ago. Originally, feature-based algorithms were applied to extract feature points and then search the corresponding features in the overlapping images. The restriction to matches of selected points usually provides correspondences at high certainty. However, feature-based matching was also introduced to avoid problems due to limited computational resources. In contrast, recent stereo algorithms aim at dense, pixel-wise matches. By these means, 3-D point clouds and digital surface models (DSM) are generated at a resolution that corresponds to the ground sampling distance GSD of the original images. To compute pixel matches even for regions with very limited texture, additional constraints are required. Local or window-based algorithms like correlation use an implicit assumption of surface smoothness since they compute a constant parallax for a window with a certain number of pixels. Those local algorithms establish references between images only under consideration of the gray-value properties of a small environment. This may be error prone, because small variations of the gray values and repetitive patterns are difficult to control. In contrast, so-called global algorithms use an explicit formulation of this smoothness assumption, which is then solved as a global optimization problem [54]. In a way, these global algorithms allow a comparison of the results of the local computation and, thus, allow for a detection of mismatches and the subsequent deletion of the outliers. One example is scanline optimization, which can be solved very efficiently by recursive algorithms. Scanline optimization is applied where beforehand the local image analysis has been done row by row with mostly a not fully fitting edge detection. The resulting image looks frayed. Using the scanline optimization, the edge points are averages with a geometrically proper and nicely looking result. A very popular and well-performing example is semiglobal matching [55], which evaluates a cumulative cost function from the scanlines in the four cardinal and four ordinal directions, east, northeast, north etc., Though the algorithm operates in two dimensions, it is still fast, because it substitutes the 2-D computation by 8 1-D computations. Especially when it is combined with sophisticated aggregation strategies it can produce accurate results very efficiently [54].

The progress of software tools for image-based DSM generation is also documented by a benchmark conducted by the European Spatial Data Research Organization (EuroSDR) [56]. This benchmark tested the DSM programs listed in Table 9.9.

In addition to the generation of 2.5-D models like DSM and DTM, the extraction of real 3-D structures especially in dense and complex urban environments is becoming increasingly important. Such meshed surface representations are widely used for visualization purposes. Moreover, since they directly represent neighborhood information they are especially useful in follow-up processes aiming at semantic interpretation of 3-D data.

9.4 Sensor Fusion

9.4.1 Principle

In Sect. 9.3, we described how to reconstruct a 3-D scene using structure-from-motion techniques. Such methods rely solely on image observations to determine the relative orientation between images at an arbitrary scale (e.g., 1). Other observations need to be added to resolve the scale correctly and to obtain the coordinates of objects in a reference/mapping frame. Ideally, such additional information is fused together with image observations in such a way that allows

9

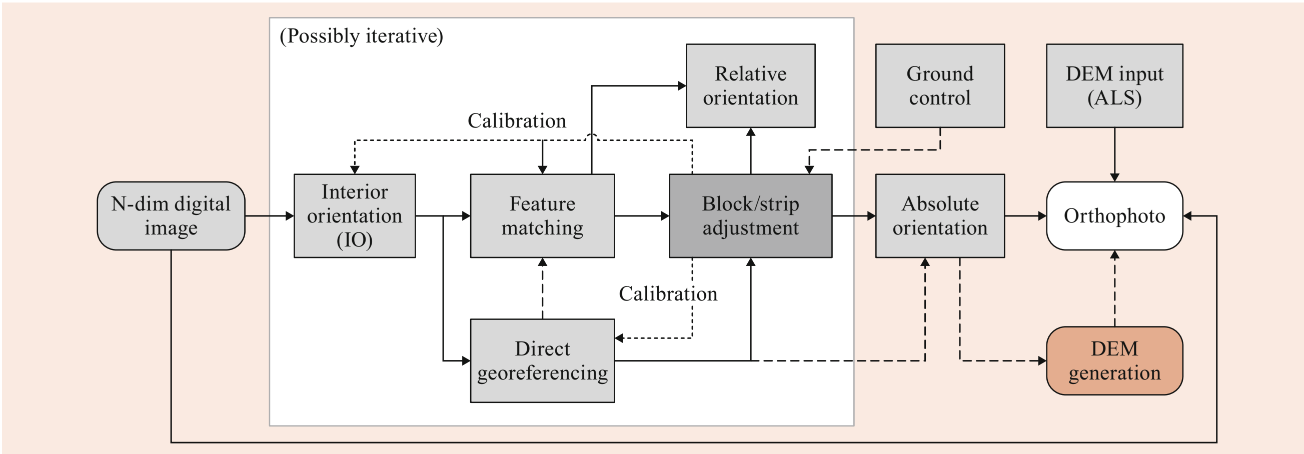


Fig. 9.32 General overview of sensor fusion, with the optimization step denoted as block/strip adjustment, also known as assisted aerotriangulation (AAT)

an optimal recovery of all parameters involved, including those related to the unknown parameters of optical sensors for the purpose of mapping. This, generally, requires determining:

1. The sensors' interior orientation (IO) parameters. In the case of frame/line imagery, a basic set of IO parameters may comprise the focal length (f) or principal distance (c), the principal point (x_0, y_0) and lens distortions that allow consistent interpretation of sensor data (in the case of lidars this may be rangefinder bias or misalignment between the laser-beam reflecting mirror and its angular encoder).
2. The sensors' absolute exterior orientation (EO) that, similarly to relative orientation, geometrically connects the sensor data among them, plus with respect to the real world (Sect. 9.2) (this is also referred to as pose when limited to position and attitude).
3. Auxiliary system parameters that geometrically relate different sensors with respect to each other in space and time, such as lever arms, bore-sights, or time-stamping offsets.

As depicted in Fig. 9.32, the process of sensor orientation may take different paths (dashed versus full lines):

- The sensor data can be oriented directly using navigation technology (Sect. 9.2).
- The sensor is oriented indirectly by identifying corresponding features across overlapping parts of data and connecting them with external references on the ground (Sect. 9.3.7). This is achieved by a procedure that will be referred to as bundle block (or strip) adjustment. In the context of airborne mapping with passive imagery, this approach is called aerial triangulation (AT).

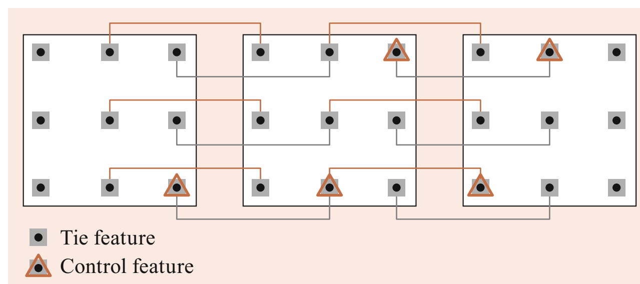


Fig. 9.33 Feature (point) conditioning on overlapping imagery

- The methods of direct and indirect sensor orientation can be combined by extending the adjustment input to a block/strip for navigation data. In this case, the procedure is called integrated sensor orientation and can be considered as an extension of the aforementioned block adjustment/AT that is referred to as assisted aerotriangulation (AAT). In robotics, this is operated sequentially and possibly in real time, the reason why it is called simultaneous localization and mapping (SLAM).

Aerial triangulation or SLAM belong to the category of network adjustment techniques that make use of redundant information in the overlapping parts of optical data (either pair-wise, strip-wise, or block-wise; Fig. 9.33). The overlapping segments are called homologous features and range from geometrical primitives as points or lines to more complex features such as surfaces. These features are conditioned within the sensor models to take the same coordinates as in the object coordinate system. This approach is applicable to passive (Sect. 9.1.1) as well as active (Sect. 9.1.3) optical sensors and is indispensable for calibration purposes.

The navigation data usually enter the adjustment as absolute or relative poses, as shown in the upper part of Fig. 9.34.

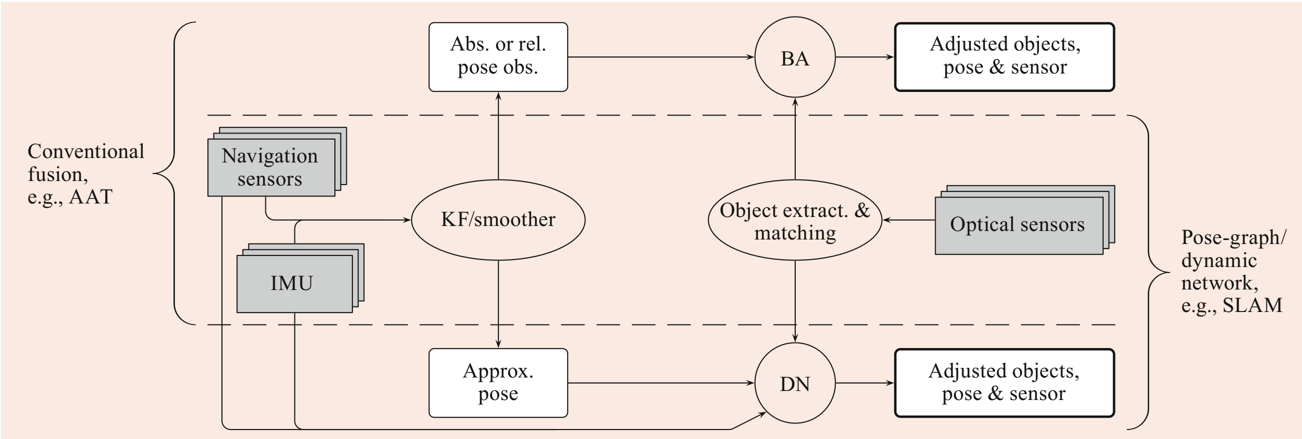


Fig. 9.34 Comparison of traditional and modern methods for fusing data from optical and navigation sensors in mapping

Table 9.10 Example of parameters for sensor orientation and calibration

Sensor	Exterior	Interior and system parameters
Frame cameras	Position, attitude, (constant trajectory errors per block or strip)	Lens distortions, PP, PD, (temperature, pressure, bore sight, lever-arm, synchronization, etc.)
Line cameras	Position, attitude, (constant trajectory errors per block or strip)	Lens distortions, PP, PD, (temperature, pressure, bore sight, lever-arm, synchronization, etc.)
Lidar	Position, attitude, (constant trajectory errors per block or strip)	Bore sight (rangefinder offset, mirror distortion & alignment, encoder scale & offset, etc.)
RADAR	Position, attitude, velocity	Doppler, bore sight, lever-arm, (synchronization), etc.
High-orbit satellites	Polynomial parameters	(Sensor dependent)
Low-orbit satellites	Position, attitude (or polynomial parameters)	(Sensor dependent)

If satellite positioning is absent (indoor environment), intermittent (terrestrial vehicles), or the inertial observations are of poor quality (small UAVs, mobile robots), the trajectory determination process may result in time-varying biases in position and attitude, the character of which cannot be correctly modeled within the network of this type. In such a situation it is better to introduce the original inertial readings (i.e., angular rates and specific forces) directly into a modified network, as depicted in the lower part of Fig. 9.34. This approach is rigorous but requires some special care when introducing the differential equations relating the inertial observation to poses. This method proposed by [57] under the term dynamic network bears a number of advantages as well as challenges. With a number of simplifications that are not appropriate for airborne mapping, this approach is also employed in robotic indoor SLAM where it is referred to as pose-graph estimation [59]. Its extended modeling, which is applicable to large-scale mapping project, is described in [60].

9.4.2 Parameters

The goal of setting up a network is the optimal fusion of all sensor data to determine concurrently and optimally the coordinates of the image features in the mapping frame together with the set of orientation parameters (exterior, interior, system). An example of parameters sets for different sensors is presented in Table 9.10. More specifically, in the example of frame/line cameras, these unknown parameters are:

- 3-D positions of distinctive features identified in the images (e.g., tie points), $\boldsymbol{p}_n^m$, with $n \in \{1; \ldots; N\}$
- (optionally) basic interior orientation parameters, i.e., the camera constant c, the principal point (x_0, y_0)
- (optionally) additional interior orientation parameters represented either by physical or the replacement models (Sect. 9.4.3)
- samples of the IMU-body frame pose for each camera exposure j, i.e., the position and the attitude of body frame with respect to mapping frame m; $\Gamma_{b,j}^m = [\boldsymbol{x}_b^m(j), \mathbf{R}_b^m(j)]$, with $j \in \{1; \ldots; J\}$
- (optionally) the body to camera lever-arm and bore sight $\boldsymbol{x}_{bs}^b$ and $\mathbf{R}_s^b$
- (optionally in dynamic network/SLAM or in AAT if GNSS position is used instead of GNSS/INS) GNSS antenna lever-arm $\boldsymbol{x}_{ba}^b$, or $\boldsymbol{x}_{sa}^s$, respectively
- (optionally in dynamic network/SLAM) INS systematic errors, e.g., random, yet time-constant 3-D bias vectors for the gyroscopes $\boldsymbol{b}_\omega$ and accelerometers $\boldsymbol{b}_f$.

The observation models related to the interior orientation of some optical sensors and those related to navigation sensors are presented in the following.

9.4.3 Optical Distortion Models

Optical distortions directly influence the metric quality of the image and, therefore, have to be considered. As introduced in Sect. 9.1 for the case of an ideal camera where the incident and emerging nodal points define its optical axes, the chief or central rays pass through the lens without deviation, while the emerging ray remains parallel to the original incident ray. The deviation from this ideal, parallel, case needs to be modeled (and later estimated by the calibration setup) since the ideal assumptions cannot be perfectly met in a real camera system design. Such deviations can be best captured by models that relate to the physical properties of the system. When this is not possible, either due to the unknown properties of the systems or its high complexity, it may be better to adopt some general models (e.g., polynomial) and determine a subset of relevant parameters.

Sensor Physical Models

In many frame/line cameras, symmetric lens distortion has the most relevant influence on 3-D object point reconstruction. Relation (9.33) introduced a basic distortion model of perspective-centered image coordinates. A more general model is the Conrady–Brown distortion correction [61] relating the distorted image coordinates $(x_\mathrm{d}, y_\mathrm{d})$ to the undistorted (x, y) through third-order radial $[k_1, k_2, k_3]$ and second-order tangential $[p_1, p_2]$ coefficients

$$\begin{aligned} x_\mathrm{d} &= x\left(k_1 r^2 + k_2 r^4 + k_3 r^6\right) + p_1\left(r^2 + 2x^2\right) + 2p_2 xy \;, \\ y_\mathrm{d} &= y\left(k_1 r^2 + k_2 r^4 + k_3 r^6\right) + p_2\left(r^2 + 2y^2\right) + 2p_1 xy \;, \end{aligned} \tag{9.70}$$

with $r^2 = x^2 + y^2$. Affinity as well as nonorthogonality effects on image coordinates maybe added to (9.70), and the size of a particular calibration set may be even larger depending on the system and the type of calibration [62]. When radial distortion is present, the image point is displaced radially in comparison with its ideal position. If this displacement is positive, i.e., the point is shifted towards the image borders, the distortion is referred to as barrel distortion; if the distortion is negative, it is referred to as pincushion distortion. Radial distortions can be balanced through proper adaptation of the focal length f. As can be seen from Fig. 9.35, the distortion remaining after balancing is small. Since this modified value is an outcome of the camera calibration and different from the physical focal length, it is now called the calibrated focal length or camera constant c.

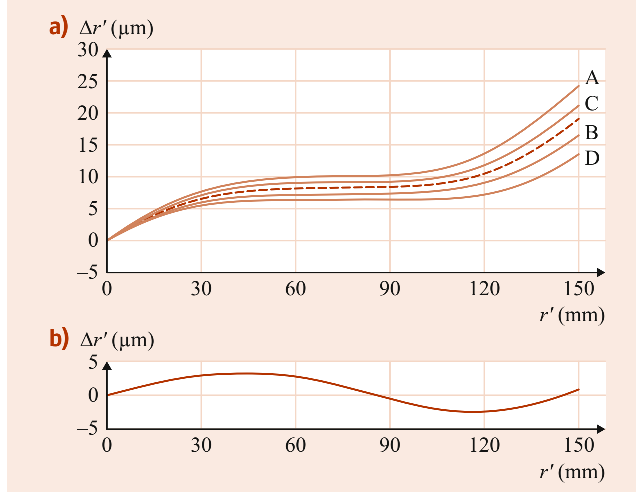


Fig. 9.35 (**a**) Radial distortion curves (A–D) and their mean value (*dashed line*). (**b**) Mean radial distortion after balancing, after [2]

Sensor Replacement Models

A sensor replacement model is a model that approximates the original, or rigorous, sensor model associated with a specific sensor by an arbitrary function. Although such models hide the details of the physical sensor model, they have some advantages, possibly being applicable across different sensors. Also, their evaluation may be faster for obtaining ground-to-image coordinates. This is especially interesting for voluminous satellite data or real-time mapping applications. Examples of replacement sensor models include:

- The 3-D polynomial model
- The affine line-based transformation model (e.g., satellite imagery)
- The rational polynomial coefficients model (e.g., satellite imagery)
- The universal sensor model (e.g., in general image-processing packages).

Such replacement models are successfully applied for transfer of sensor orientation to the final users; however, their usage in block adjustment is less appropriate.

9.4.4 Observation Models

Image Observations

Frame cameras

Here we combine the basic camera model (9.32) that gives the undistorted image coordinates of a mapped feature and the geometrical relations between the optical and navigation data (Fig. 9.23 and Sect. 9.2.5). Let $\boldsymbol{p}_n^m$ be the 3-D coordinates of the n-th tie-point expressed in the mapping frame. Considering the poses of the b-frame associated with the j-image $\Gamma^m_{b,j} = [\boldsymbol{x}^m_b(j), \mathbf{R}^m_b(j)]$, through the camera bore sight $\mathbf{R}^b_s$ and lever-arm $\boldsymbol{x}^b_{bs}$, this point is projected to the image plane of an ideal perspective camera with camera constant c and principal point (x_0, y_0) at image coordinates (x, y), such that

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \frac{1}{\mu}\begin{pmatrix} c & 0 & x_0 \\ 0 & c & y_0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \left[\left(\mathbf{R}^m_b(j)\mathbf{R}^b_s\right)^\mathsf{T}\left(\boldsymbol{p}^m_n - \boldsymbol{x}^m_b(j)\right) - \left(\mathbf{R}^b_s\right)^\mathsf{T}\boldsymbol{x}^b_{bs}\right]. \tag{9.71}$$

The scale factor μ can be eliminated by rearranging the equation system (9.71) so that the image coordinates are separated on the left-hand side and then dividing the first two relations. Then, the distorted image coordinates $(x_\mathrm{d}, y_\mathrm{d})$ can be determined, e.g., as in (9.70). Finally, with $\boldsymbol{l}_{n,i}$ being the image coordinate of the n-th 3-D point on the image, as reported by the tie-points detection algorithm, the image observation model reads

$$\boldsymbol{l}_{n,i} + \boldsymbol{v}_{n,i} = \begin{pmatrix} x_\mathrm{d} \\ y_\mathrm{d} \end{pmatrix}, \tag{9.72}$$

where $\boldsymbol{v}_{n,i}$ is the correction vector.

Line cameras

The collinearity condition expressed for frame cameras (9.71) with the column (x) and row (y) image coordinates can be adapted to line cameras by omitting the y or row pixel coordinates

$$\begin{pmatrix} x \\ 0 \\ 1 \end{pmatrix} = \frac{1}{\mu}\begin{pmatrix} c & 0 & x_0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \left[\left(\mathbf{R}^m_b(j)\mathbf{R}^b_s\right)^\mathsf{T}\left(\boldsymbol{p}^m_n - \boldsymbol{x}^m_b(j)\right) - \left(\mathbf{R}^b_s\right)^\mathsf{T}\boldsymbol{x}^b_{bs}\right]. \tag{9.73}$$

Nevertheless, most line cameras are multiple m-line cameras (with lines $k = 1, \ldots, m$). In such a case, with the use of projection matrix $\mathbf{K}$ (9.31), the undistorted collinearity model reads

$$\begin{pmatrix} x_k \\ 0 \\ 1 \end{pmatrix} = \frac{1}{\mu}\mathbf{K}\Big[\left(\mathbf{R}^m_b(j)\mathbf{R}^b_p\mathbf{R}^p_{s,k}\right)^\mathsf{T}\left(\boldsymbol{p}^m_n - \boldsymbol{x}^m_b(j)\right) - \left(\mathbf{R}^b_p\right)^\mathsf{T}\boldsymbol{x}^b_p + \boldsymbol{x}^p_{bs,k}\Big], \tag{9.74}$$

where p is the common camera-platform reference frame, $\boldsymbol{x}^p_{bs,k}$ is the lever-arm of line k to p-frame origin, and $\mathbf{R}^p_{s,k}$ is the rotation from the k-camera to the platform frame. With $\boldsymbol{l}^k_{n,i}$ being the distorted image coordinate of the n-th 3-D point on the image, as reported by the tie-points detection algorithm for the line k, the image observation model reads

$$\boldsymbol{l}_{m,n,i} + \boldsymbol{v}_{m,n,i} = \boldsymbol{x}_{d,m}\,. \tag{9.75}$$

The relation between distorted x_k and undistorted image coordinates $x_{d,k}$ depends on the optical model described in Sect. 9.4.3, e.g., the first line of (9.70).

Ground Control

If available, observations of the 3-D coordinates of the n-th mapped feature are introduced as

$$\boldsymbol{l}_n + \boldsymbol{v}_n = \boldsymbol{p}^m_n\,. \tag{9.76}$$

Position

The position of the sensor is determined either by a GNSS receiver or by GNSS/INS integration. The first observation refers to the antenna phase center a, the latter is usually the origin of the b-frame. Both are determined with respect to some global coordinate system, e.g., WGS-84, which can be transformed to the m-frame. Referring to (9.14) and considering the positions $\boldsymbol{x}^m_{a/b}$ with respect to m together with the lever-arm $\boldsymbol{x}^{a/b}_s$ and attitude $\mathbf{R}^m_b$, the observation model for sensor position reads

$$\boldsymbol{l}_{p,j} + \boldsymbol{v}_{p,j} = \boldsymbol{x}^m_{a/b}(j) + \mathbf{R}^m_b(j)\boldsymbol{x}^{a/b}_s\,. \tag{9.77}$$

When the chosen mapping frame is a projection and the sensor-fusion model is derived for a Cartesian frame, the position observation should be corrected in height. As described in [30], the size of such a correction depends on the absolute terrain height, the flying altitude above it, and the value of the projection scale at the perspective center.

Velocity

Velocity observations are needed for sensors like radars. They can be also useful for estimating a time-stamping offset between optical and navigation data [63]. GNSS or GNSS/INS provide velocity observation of the antenna a or body b, either in e or l frames, respectively. Similarly to position, the velocity vector can be transformed to a mapping frame. The velocity observation model is then

$$\boldsymbol{l}_v + \boldsymbol{v}_v = \boldsymbol{v}^m_{a/b}(j) + \mathbf{R}^m_b(j)\boldsymbol{\Omega}^b_{mb}(j)\boldsymbol{x}^{a/b}_s\,, \tag{9.78}$$

where $\boldsymbol{\Omega}^b_{mb}(j)$ is the skew-symmetric matrix of angular velocity vector $(\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3) = (\Omega_{3,2} = -\Omega_{2,3}, \Omega_{1,3} = -\Omega_{3,1}, \Omega_{2,1} = -\Omega_{1,2}))$ between the m and b-frames expressed in the b-frame.

9

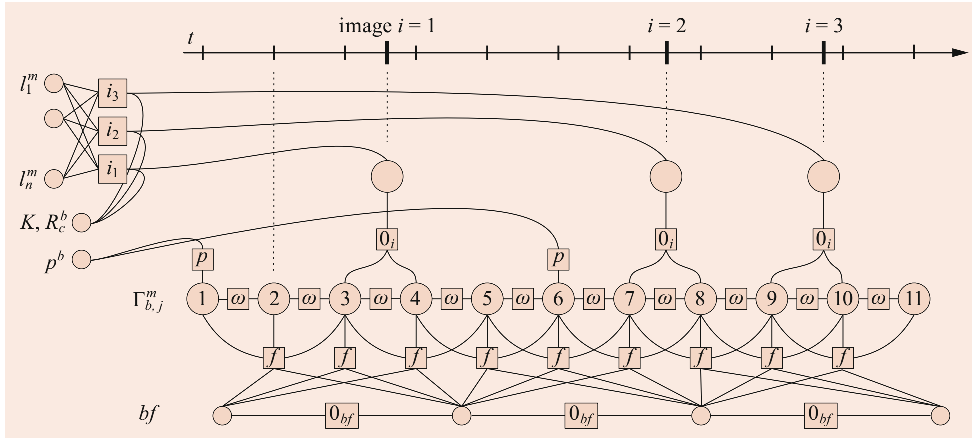


Fig. 9.36 A simplified instance of a dynamic network formulated as a factor graph

Attitude

Corrections to attitude observations $\boldsymbol{v}_R$ are expressed as non-commutative multiplication of rotation matrix $\mathbf{R}_v$. Hence, if attitude is observed externally as $\mathbf{R}_m^b$, the attitude observation equations reads

$$\boldsymbol{l}_R + \boldsymbol{v}_R \equiv \mathbf{R}_m^b\,\mathbf{R}_v = \mathbf{R}_s^b\,\mathbf{R}_m^s\;, \tag{9.79}$$

with $\mathbf{R}_s^b$ being the bore sight and $\mathbf{R}_m^s$ the mapping-to-sensor frame rotation. However, INS/GNSS processing usually delivers $\mathbf{R}_l^b$, where the orientation of the local-level frame l with respect to the m-frame changes with the change of position. In such a case, the attitude observation equations need to be modified to

$$\mathbf{R}_l^b\,\mathbf{R}_v = \mathbf{R}_s^b\,\mathbf{R}_m^s\,\mathbf{R}_l^m\;, \tag{9.80}$$

where the definition of $\mathbf{R}_l^m$ depends on the arbitrary choice of the mapping frame. For instance, with the mapping frame defined as the Cartesian system on a tangent plane of a WGS-84 ellipsoid at geographical coordinates (ϕ_0, λ_0)

$$\mathbf{R}_l^{m\equiv l_0} = \mathbf{R}_e^l(\phi_0, \lambda_0)\,\mathbf{R}_l^e(\phi_t, \lambda_t)\;, \tag{9.81}$$

where $\mathbf{R}_l^e = (\mathbf{R}_e^l)^\mathsf{T}$ is defined by (9.11).

Conformal projections are often used when mapping in country-specific national coordinates. There, the convergence of meridian γ_{PC} at each perspective-center position (PC) needs to be accounted for. Considering the east-north-up axis convention usually used in projections, the attitude observation for the projection reads

$$\mathbf{R}_l^{m\equiv p} = \mathbf{R}_{\mathrm{NED}}^{\mathrm{ENU}}\,\mathbf{R}_3(\gamma_{\mathrm{PC}})\;, \tag{9.82}$$

where the matrix on the left-hand side involves exchanging the first and second axis and mirroring the third one, and the second term is a standard rotation matrix about the third axis of the (modified) p-frame with the meridian convergence value γ_{PC} computed at the sensor perspective-center for the particular projection. When the mapping frame is a conformal projection defined on a national reference ellipsoid, the observation equation for attitude may need to be further modified for the relative rotation between the ellipsoid employed for INS/GNSS integration and that of the national datum [31].

Angular Velocity and Specific Forces

As outlined in Sect. 9.4.1, some method of robotic's SLAM or dynamic networks, directly employs the inertial raw observations, i.e., the angular velocities ω and specific forces f. The rigorous form of these observation equations is rather long and is described in detail in [60]. As schematically depicted in the bottom part of Fig. 9.36, ω and f constrain the unknown poses $\Gamma_{b,j}^m$ via differential relations that are approximated by first and second-order finite differences. Also shown schematically in Fig. 9.36 are the connections between other previously mentioned observation models (represented by boxes) with the unknown parameters (represented by circles): p for GNSS positions, l_n^m for image observations, and 0 for so-called zero observations. The latter relates some parameters by known functional relationships without being associated with an actual sensor reading, e.g., interpolation between poses to image observation times 0_i, or time-correlated evolution of accelerometer biases 0_{bf}. Although usually applied, the evolution of gyroscope models $0_{b\omega}$ is not represented in Fig. 9.36 for the sake of clarity.

9.4.5 Estimation

The goal of estimation is in the optimal combination of all observations that leads to the most correct values of unknown parameters. Assuming that the correction or residual vectors $\boldsymbol{v}$ of all observations are randomly distributed, this is achieved by solving a nonlinear weighted least-squares problem. Gathering all the terms on one side of each observation equation will result in the following condition

$$g(\boldsymbol{l}+\boldsymbol{v},\boldsymbol{x})=0\,, \tag{9.83}$$

where $\boldsymbol{l}$ represents the vector of the given observation, $\boldsymbol{x}$ is the vector of unknown parameters, and $\boldsymbol{v}$ is the vector of residuals of the observations that are assumed to be normally distributed, i.e., $\boldsymbol{v} \approx N(0, \mathbf{C}_{\ell\ell})$. Although the conditioning relation $g(\cdot)$ varies per sensor and observation, the general estimation methodology by the least-squares principle stays the same:

- First, a linear model is obtained by linearizing the (nonlinear) function $g(\cdot)$ according to the observations $\boldsymbol{l}$ and parameters $\boldsymbol{x}_i$, where the index $i=0$ denotes its initial approximation

$$g(\boldsymbol{l},\boldsymbol{x}_i)+\left(\frac{\partial g}{\partial \boldsymbol{l}}\right)_{|(\boldsymbol{l},\boldsymbol{x}_i)} v+\left(\frac{\partial g}{\partial \boldsymbol{x}}\right)_{|(\boldsymbol{l},\boldsymbol{x}_i)}\Delta\,\boldsymbol{x}_i=0 \tag{9.84}$$

  or

$$-\,\boldsymbol{g}_i-\mathbf{B}_i\,\boldsymbol{v}+\mathbf{A}_i\,\Delta\,\boldsymbol{x}_i=0\,, \tag{9.85}$$

  where $\boldsymbol{g}_i=-g(\boldsymbol{l},\boldsymbol{x}_i)$, $\mathbf{B}_i=-(\partial g/\partial \boldsymbol{l})_{|(\boldsymbol{l},\boldsymbol{x}_i)}$, and $\mathbf{A}_i=(\partial g/\partial \boldsymbol{x})_{|(\boldsymbol{l},\boldsymbol{x}_i)}$.
  (In the case of $\boldsymbol{v}=-\boldsymbol{l}+g(\cdot)$, this derivative is trivial.)
- Second, the nonlinear model (9.83) is solved by iterating the solutions of the linear model (9.85) to convergence. The corrections to the parameters Δx_{i+1} are obtained by the best linear unbiased estimation (BLUE). As for the parameters, the BLUE estimation is the solution of the so-called normal equation [64]

$$\Delta\boldsymbol{x}_i=\left(\mathbf{A}^\mathsf{T}(\mathbf{B}\mathbf{C}_{vv}\mathbf{B}^\mathsf{T})^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^\mathsf{T}(\mathbf{B}\,\mathbf{C}_{\ell\ell}\mathbf{B}^\mathsf{T})^{-1}\boldsymbol{g}_i\,, \tag{9.86}$$

  while the corrections to observations are obtained as

$$\boldsymbol{v}_i=\mathbf{C}_{vv}\mathbf{B}^\mathsf{T}(\mathbf{B}\mathbf{C}_{\ell\ell}\mathbf{B}^\mathsf{T})^{-1}(\mathbf{A}\boldsymbol{x}-\boldsymbol{g}_i)\,. \tag{9.87}$$

After each iteration, the set of parameters is rectified by the estimated correction $\Delta\boldsymbol{x}_i$ as $\boldsymbol{x}_{i+1}=\boldsymbol{x}_i+\Delta\,\boldsymbol{x}_i$, and the observations are updated according to $\boldsymbol{l}_{i+1}=\boldsymbol{l}_i+\boldsymbol{v}_i$. The linearization of (9.84) is repeated with the updated set of parameters and observations ($x_{i=0}$ and $l_{i=0}$ denotes parameters and observation initial values, respectively). The iteration is stopped when the corrections to parameters $\Delta\,\boldsymbol{x}_i$ are no longer significant (i.e., $\Delta\,\boldsymbol{x}_i\approx 0$). After convergence, the last iteration step is repeated with the original observations ($\boldsymbol{l}_{i=0}$). The respective covariances characterizing the accuracy of parameters and measurements are estimated in parallel at each step by the relations presented in [64].

The quality of the estimation may be judged according to the analyses of residuals and global a-posteriori estimation of the variance. The latter is evaluated as $\widehat{\sigma}_0{}^2=(\boldsymbol{v}^\mathsf{T}\,\mathbf{C}_{\ell\ell}{}^{-1}\,\boldsymbol{v})/(n-u)$, where n is the number of observations and u the number of parameters. Special situations may lead to some simplification of the general model (9.84) and its solution (9.86). Detailed information on this subject is presented in [65] and [64].

The general formulation of the sensor fusion may be very large, leading to hundreds of thousands unknowns (or even millions of unknowns for dynamic networks) but is inherently sparse and can be solved efficiently exploiting state-of-the-art least-squares solvers [66] based on very efficient sparse linear algebra routines [67]. (Reducing the number of unknowns in dynamic network is possible by preintegrating a certain number of IMU observations [68].)

9.4.6 Adopted Approaches

Frame Sensors

Advances in computer vision and digital-image processing enabled fully automated selection and measurement of corresponding points, which together with satellite positioning improved the productivity and accuracy of mapping. The stability of GNSS-assisted aerotriangulation remains dependent on the image texture, whose variation may cause problems in large-scale or oblique imagery, in forested areas, or over snow-covered landscape. These problems can be somewhat mitigated with the concurrent employment of integrated inertial navigation that allows also direct orientation. The latter concept found its place in fast mapping, applications of lower mapping accuracy, corridor mapping, and terrestrial mobile mapping. Absolute orientation based solely on ground control points remains in use for small mapping projects, such as those performed by small drones without RTK capacity.

GNSSs are included for quality control or for calibration purposes on larger missions, benefiting navigation technology. Indeed, when factors such as accuracy and reliability are important, the method of integrated sensor orientation remains the most sophisticated alternative for frame-camera orientation (Table 9.11). In this method, the first approximation of exterior orientation is provided by the navigation technology, which is present on all modern large-scale digital cameras and autopiloted drones (for the purpose of navigation, guidance, and control capacity on automated missions). Knowledge of the initial EO limits the search space for homologous points and thus improves their transfer between images on challenging texture. In this regard, external knowledge of attitude parameters is more important for oblique photography or situations with corresponding image texture than for vertical configurations and where sufficient image texture is available. In the next step, the optimization is run first to eliminate outlier observations and later to provide the final solution to the orientation problem. Ground control points are included for quality control or for calibration pur-

Table 9.11 Comparison of main orientation approaches

Aerial Observations	Advantages	Disadvantages
None	• Independent of airborne satellite signal • Simple processing chain • Independent of navigation quality • Independent of synchronization errors	• Impractical over large or steep areas • May lead to systematic deformations (func. of GCPs) • IO correlated to EO • Large overlaps required
Position	• Absorbs IO instability • Consistent determination of all parameters • Potential for radiometric adjustment • Self-calibrating and possibly no GCPs	• Weak geometry at block ends • Not ideal for corridors • Larger side-overlap required • Textureless (e.g., costal) mapping is difficult • Problematic transfer of points in oblique imagery
Full	• Suitable for corridors & multisensor systems • ≈20% side-overlap OK • Automation, no GCPs	• Lower redundancy in corridors • Attitude accuracy dependent

Table 9.12 Indicative frequency of sensor deployment and orientation method used for frame cameras across different platforms: (–) = never, (+) = rare, (++) = sometimes, (+++) = common

		Satellite	Aircraft	Vehicle
On-board sensors	GNSS	++	+++	+++
	IMU	++	++	+++
	Star Tracker	+++	+	–
External measurements	GCP	+++	++	+
Orientation approach	Direct	+	++	+++
	AAT/(GNSS)	++	+++	+
	Integrated	+	++	+

poses. Similarly to position-assisted AT, the use of full aerial control results in lower correlation between EO/IO parameters. Table 9.12 indicates common orientation approaches across different platforms. Generally, it is acknowledged that object–space accuracy is two to four times better when using an integrated rather than a direct sensor orientation approach.

Line Sensors

Theoretically, the orientation of line sensor data can also be determined indirectly, in a manner similar to frame imagery [69]. However, this approach is rarely used in practice because the computational effort is large, and the resulting mapping accuracy is lower without the support of attitude and position sensors [70]. Also, to ensure sufficient overlap between successive exposures, the line-camera head needs to be placed on a stabilized mount. Such stabilization can be more precise when based on a real-time GNSS/INS trajectory, which is also the case for modern line cameras.

The common approach to line-camera orientation is depicted in Fig. 9.37. The onboard GNSS/INS measurements are recorded for postprocessing (PPK) and integration. At the same time, a real-time navigation solution based on point-positioning GNSS/INS integration is used to steer the camera platform stabilization. The captured images are stored and rectified during postprocessing using the best available calibration parameters and the improved EO parameters from the postprocessed trajectory. The distortion in the original imagery, caused by the motion of the sensor, is removed by this rectification, and the resulting scenes can be viewed stereoscopically. The automated matching process is performed, but the tie-points coordinates are referred back to the original imagery. The orientation parameters are updated by the block adjustment using image measurements and orientation parameters. Possibly, ground control points (GCPs) may also be included for calibration, improved accuracy, or for control purposes. The block adjustment provides final orientation parameters that are applied either to the prerectified images for DEM (digital elevation model) generation or also directly to the raw images for the (best possible) orthophoto production.

Investigations with modern line cameras have revealed that when a real-time GNSS/INS solution is used to support the matching process, the number of matched points is approximately 25% less compared with the use of a postprocessed trajectory [70]. Also, the accuracy of object points was two to three times worse when using EO parameters based on a real-time trajectory, which is not acceptable for applications with the highest accuracy demands (i.e., ≤ 0.1 m in object space). It was also observed that the mapping accuracy is two to four times worse when based on direct orientation as compared to integrated sensor orientation.

Calibration

As soon as sensors such as GNSS are added to derive the camera perspective-center coordinates in order to directly measure the exterior orientation elements with high absolute accuracy, systematic differences between perspective center coordinates derived from a bundle adjustment without "assistance" and directly measured coordinates are likely to occur. These differences cannot always be attributed to errors in the trajectory computations, especially when GNSS trajectory solutions can be delivered with high accuracy. Thus, such differences might also be caused by changes in the cam-

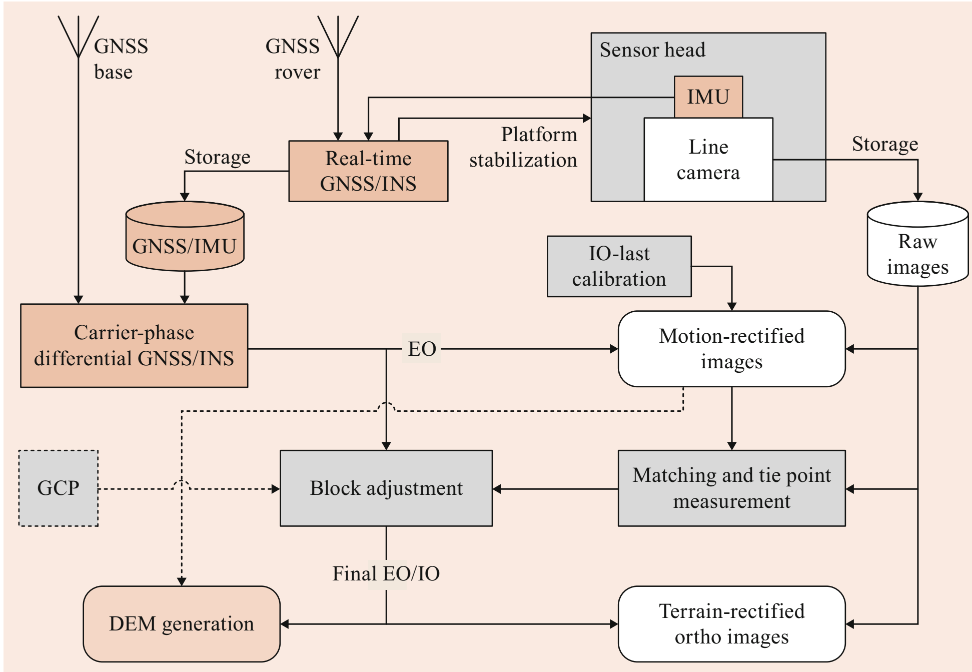


Fig. 9.37 Processing chain for line-camera data

era geometry. For example, since airborne images are mostly taken in nadir direction, incorrect assumptions of the focal length will shift the adjusted perspective center coordinates along the vertical axis. This immediately causes offsets between the directly measured perspective center coordinates and the coordinates obtained from the photogrammetric bundle adjustment. On the other hand, the availability of direct exterior orientation measurements of sufficiently high accuracy now offers the possibility to completely calibrate the camera geometry based on in-situ approaches even for airborne sensors. However, a flat field such as the Earth's surface combined with parallel viewing directions does not allow for determination of the full camera geometry, unless the test field is of special design (e.g., it has significant height variations). Such requirements are necessary to suppress the high correlations between unknown exterior orientation elements and certain sensor parameters (i.e., sensor interior orientation). The additional sensor parameters are estimated in extended aerial triangulation. Additional parameter models that directly relate to physical changes in the sensor geometry are well established in photogrammetric imaging [64] but have rarely been used for airborne camera calibrations in the past. Due to the previously described correlations, mostly mathematical polynomials were preferred to overcome remaining systematic effects in airborne imagery. Such parameter sets have been proposed by [71] and others. These additional parameters are not correlated with the exterior orientation elements and can, thus, be used in standard aerial triangulation, but they do not refer to changes in the camera geometry. Another aspect is the need to calibrate sensor systems instead of single system components only. This is also referred to as system versus component calibration and becomes obvious if the design of today's digital imaging sensors is considered. They typically consist of several components:

- The imaging sensor itself, which may contain several optical lens systems.
- Additional sensors for direct measurement of the sensor trajectory during data capture, which are almost standard for new digital sensors.

In contrast to film-based cameras, where calibration mainly considered the lens component only, the overall calibration of such, more complex systems cannot be done from laboratory calibrations exclusively. The relative orientations between the optical sensor and inertial measurement unit can only be derived from in-situ approaches. This method is also convenient to derive the relative positions between the GNSS antenna and the inertial and camera perspective-center.

Laser Scanners

The process of kinematic laser scanning relies on direct sensor orientation. Nevertheless, the principle of integrated sensor orientation can be introduced either for system calibration [42], the mitigation of residual systematic errors in trajectory determination [72], or both [39, 41, 73]. Such an adjustment process also serves as an internal control of the laser-scanning mission. The initial development of block adjustment in kinematic laser-scanning used the concept of tie-points (based on return-intensity values). Contrary to cameras, this principle is not very suitable, as the corre-

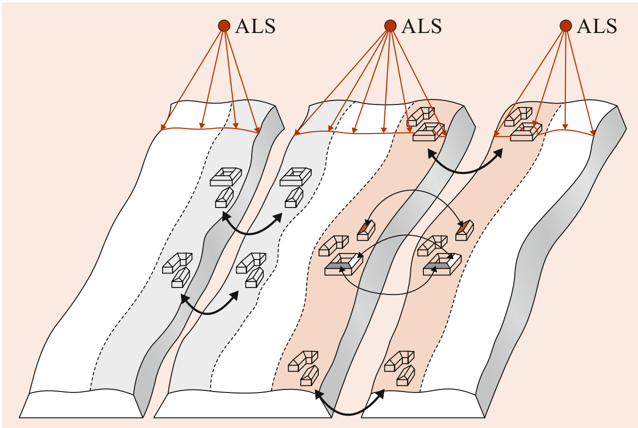


Fig. 9.38 Principle of surface-patch conditioning in airborne laser scanning. After [39]

spondence between laser points is only approximate. Modern approaches, therefore, rely on conditioning surface patches or other geometrical primitives that overlap between different passes (Fig. 9.38). The success of this approach depends on the number of patches and their form, size, and spatial variation. Generally, this approach works better on patches, whose form is known a priori. This is the case for planar surfaces on buildings or other manmade structures. The parameters of the planes are estimated together with the calibration parameters that may include also biases in the trajectory [73]. Such trajectory bias modeling is approximate and can be avoided when the estimation includes the inertial raw readings as observations [68] (Sect. 9.4.6, *Angular Velocity and Specific Forces*). When considering the simpler formulation with the platform position and orientation provided by a GNSS/INS, together with the range and encoder angle values measured by the laser scanner, there are eight observations per laser return. Using (9.15), the observation equation for a laser target in the e-frame $\boldsymbol{x}_i^e$ lying on a plane $\boldsymbol{s}_j$ is given by

$$\left\langle \boldsymbol{s}_j \,, \begin{pmatrix} \boldsymbol{x}_i^e \\ 1 \end{pmatrix} \right\rangle = 0 \,, \tag{9.88}$$

where, the plane parameters are given by

$$\boldsymbol{s}_j = (s_1 \;\; s_2 \;\; s_3 \;\; s_4)^\mathsf{T} \,, \tag{9.89}$$

with s_1, s_2, s_3 the direction cosines of the plane's normal vector and s_4 the negative orthogonal distance between the plane and the coordinate system origin. Note that the direction cosines must satisfy the unit length constraint $|\boldsymbol{s}_j| = 1$. The details of how such a constraint is added to the adjustment model are described in [42]. The principle can be extended to natural surfaces [74]; however, this approach has certain limits. First, most naturally flat surfaces are horizontal, which makes their contribution less significant. Second, perfectly flat surfaces are less common in nature, and their identification remains problematic [58]. Future methods of integrated sensor orientation with laser data will most likely start using general surface models with somewhat tighter feedback to the trajectory determination, as is the case of robotics, SLAM [59]. However, as terrestrial robots usually employ multibeam 3-D scanners, there is only one set of pose parameters per one instance of beam-array activation. This configuration is somewhat similar to that of line cameras and geometrically stronger than the more usual case of airborne scanning, where each single laser pulse has a unique set of poses.

9.5 Mapping Products

9.5.1 Surfaces

In this section, remote sensing is used in a general sense and, thus, incorporates photogrammetry. The products of remote sensing may be grouped by their production process in geometric and radiometric products. Here, we briefly introduce two typical photogrammetric products, one in 3-D and the other in 2-D: surface models and the orthophoto.

Representation

The term digital elevation model (DEM) encompasses surface representation without specifying its nature. On the other hand, the digital terrain model (DTM) is a discrete description of the physical surface (terrain), while the digital surface model (DSM) considers the terrain with all surface features, including buildings and vegetation. The information captured about terrain height by means of optical sensors (Sect. 9.1) is usually heterogeneous and unorganized. Therefore, it needs to be restructured into a form that is both comprehensive and usable for further exploitation like interpretation, visualization, manipulation, etc.

Regardless of its form, a surface model will always remain an approximation of the reality with limited resolution. However, the choice of its representation is important, as it dictates the requirements on data storage, possibility to portray sharp changes in the topography or the efficiency in model manipulation and analysis. Among the number of possibilities of terrain representation, grids, triangles, and contours are the most common and will, therefore, be discussed in more detail (Fig. 9.39).

Elevation grid

An elevation grid is the most straightforward representation of terrain (Fig. 9.39c). It is characterized by the regular, lattice organization of equally spaced points in the horizontal (x, y) projection. Each point of such a mesh contains one

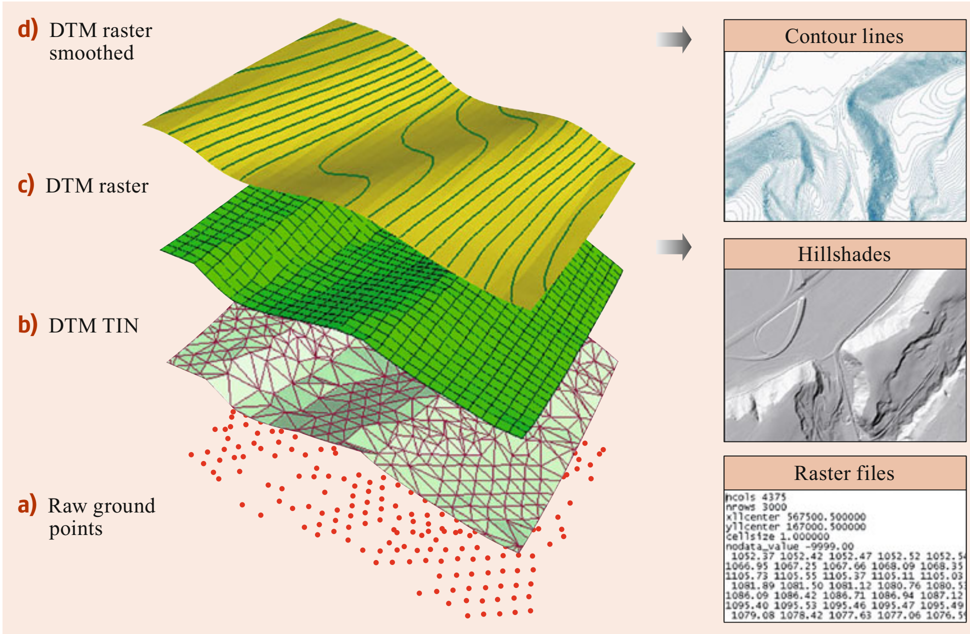


Fig. 9.39 Typical surface modeling based on irregularly distributed points (**a**). (**b**) Generation of a triangular irregular network (TIN). (**c**) Interpolation of TIN to regularly spaced grid. (**d**) Smoothing of grid and derivation of contour lines. After [12]

height (z) value for its location, which together with the x, y-coordinates is referenced to a common origin. The spacing between points is predefined and thus implies the resolution of the model. Such an organization is similar to that of an image and due to such a resemblance, this representation is referred to as a raster. In this structure, only z-values need to be stored, as the x, y-coordinates are derived from corresponding indices and cell spacing.

The grid structure is convenient for its simplicity of organization, which is also practical for further data manipulation. On the other hand, it is less suitable for the modeling of steep landscapes where the resolution with respect to slope normally decreases progressively (Table 9.13). It is also not suitable for the modeling of complex shapes in three dimensions as overhangs, because the storage of several z values per grid cell is not possible. The grid arrangement is also suboptimal in capturing characteristic landscape features like highest points or break-lines which may not coincide with a grid cell. To describe finer terrain features by this method the cell size needs to be reduced. However, this increases storage requirements without providing additional information in areas where coarser cell size is adequate. Such an inconvenience could be circumvented by allowing the cell size to be adaptive (e.g., quadtree storage) or by applying image-like compression [75].

TIN

Compared to a grid, the triangulated irregular network (TIN) structure (Fig. 9.39b) is considerably more adaptive to local terrain variations. It constitutes a set of nodes (points) that are connected by lines to form triangles. The surface within each triangle is represented by a plane facet. The individual facets fit in a mosaic that yields a surface. Such modeling is appropriate for areas with sudden changes in slope, where the edges of triangles can be aligned with discontinuities in the landscape (e.g., ridges, bottoms of gullies).

The storage requirements of irregularly spaced points and their associated TIN structure are quite large, as all three coordinates per point need to be stored separately. A solution that overcomes this problem is the above-mentioned generation of uniformly spaced elevation grids, where the x, y-coordinates are described by an array of indexes, and only the z-value is stored.

Contours

Terrain representation by contours (Fig. 9.39d) was the most common way of surface modeling before the digital era. It is also the most frequent means of coding the vertical dimension into topographical maps. Contours are lines of constant elevation (isolines) that are usually projected onto a 2-D surface. In the past, contours were generated manually from oriented photographs on stereoplotters. Although laborious, this process was accurate when carried out by a skilled operator who, at the same time, made judicious generalization of reality. Although the DTM could be derived from contours by interpolation, this practice is left to cases when a cartographic source (i.e., map) is the primary input for its generation. In modern mapping, the contours are produced automatically from a grid, TIN, or irregularly distributed elevation points [75].

A comparison between different forms of terrain representation is given in Table 9.13.

9

Table 9.13 Comparison between different forms of DEM representation; ⊕ and ⊖ denote advantages and disadvantages, respectively

	Grid	TIN	Contours
Structure	⊕ Simple	⊖ Complex	⊖ Complex
Storage	⊕ More compact	⊖ Larger	⊖ Large
Exchange	⊕ Excellent	⊖ 2.5-D possible	⊖ Difficult
Applicability	⊖ Limited 2.5-D	⊕ 3-D possible	⊖ Limited 2.5-D
Adaptability	⊖ With quad trees	⊕ Adaptable	⊕ Adaptable
Modeling	⊖ With sampling rate	⊕ Excellent	⊖ Modest
Discontinuity	⊖ Limited	⊕ Good	⊖ Limited
Operations	⊕ Fast and robust	⊖ More complex	⊖ Not practical
Usage in maps	⊕ As hillshade	⊖ Not practical	⊕ Excellent

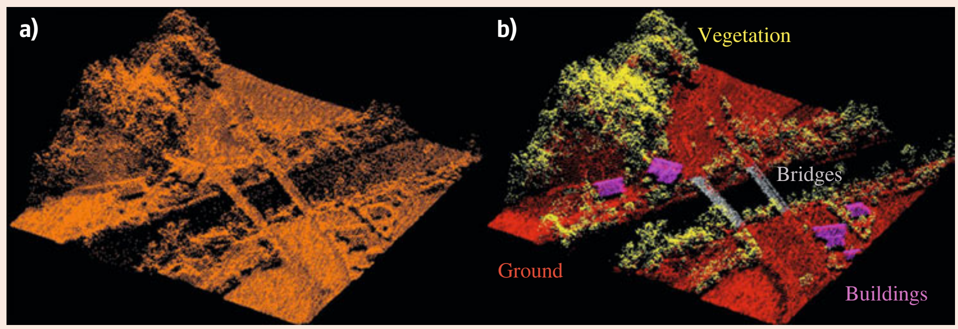


Fig. 9.40 Point-cloud classification, after [12]: (**a**) Raw laser point cloud, (**b**) point cloud classified into ground, bridge, vegetation, and building points

Reconstruction

DEMs of coarse resolution covering all continents are mostly produced by satellite missions (InSAR). The acquisition of DEMs of finer resolution at the scale of large countries is most effectively performed by airborne SAR, less effectively but more accurately by image processing using the principles presented in Sect. 9.4. Altimetric models of the highest precision and resolution usually come from airborne laser scanning.

Classification

As laser scanning is a nonselective mapping method; the acquired point cloud includes all kinds of objects (e.g., vegetation, buildings, wires) apart from the terrain itself. Hence, prior to the derivation of elevation models the point cloud needs to separated into categories of objects as depicted in Fig. 9.40. This process is called classification and is highly but not entirely automated. When the point data are obtained by insertion from oriented images (Sect. 9.4), the classification can be performed together with image matching.

Triangulation

As mentioned previously, triangulation creates a polygonal or triangular mesh from a set of unorganized points, where the facets of polygons are the discrete representations of the surface (Fig. 9.39b). Triangulation can be performed in 2-D or in 3-D, according to the geometry of input data. Large, country-like elevation models are usually created in 2.5-D, which means that the triangulation is performed in 2-D, and the z-value gets attached to each node using a unique elevation function $z = f(x, y)$. Such models are not ideal to represent steep terrain (Fig. 9.41a) or complex manmade structures. As can be seen in Fig. 9.41b, performing 3-D triangulation is more suitable for this purpose, however, its evaluation is very complex in large datasets. Also, data exchange in the GIS community is not necessarily standardized for 3-D TIN structures.

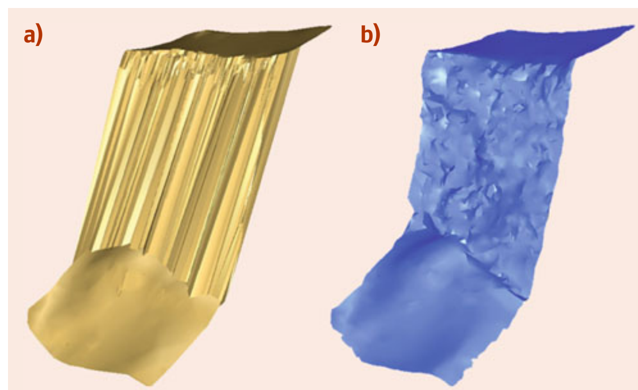


Fig. 9.41 Isometric view of a vertical rock face obtained by airborne laser scanning and triangulation; after [12]: (**a**) 2.5-D triangulation, (**b**) 3-D triangulation

Grid generation

An evenly-spaced DTM grid can be obtained by interpolation from the TIN facets. This approach is, however, only practical when the TIN model already exists. The grid can also be derived directly from the point data by various interpolation techniques. The popular interpolation methods used for this purpose are trend surface analysis, Fourier analysis, or kriging. These approaches have a global character

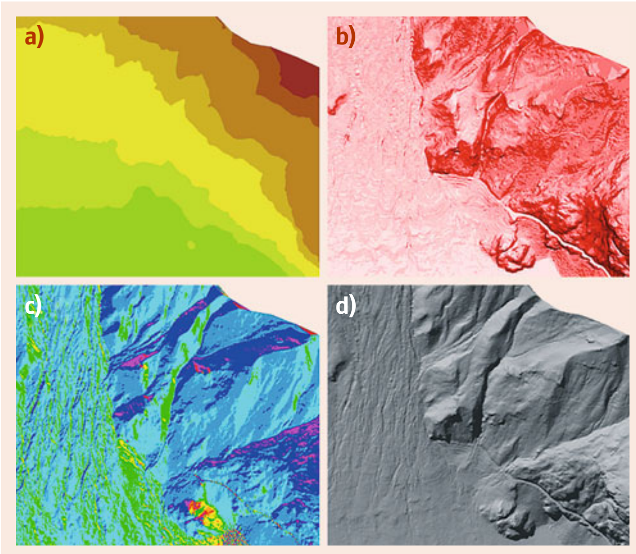


Fig. 9.42 Analysis of the raster height elevation model; after [12]: (**a**) DEM color coded by elevation, (**b**) DEM slope grid, (**c**) DEM color-coded by aspect, (**d**) DEM hillshade grid

and various levels of smoothing that are either predetermined or estimated from the data itself. On the other hand, methods of local character are more appropriate when the terrain varies abruptly, as they are based on the elevation information from the nearest points. The most frequent methods of such type are spline or cubic interpolations and inverse distance weighting [75].

In places where ground point sampling is low, like in forested areas or in dense urban zones, it is preferable to apply triangulation prior to grid generation, because the resulting DEM is less affected by the lack of data. The maximal size of data gaps to be closed by triangulation can be limited by specifying the largest length of the facet edge.

DEM Analysis

Grid representation of DEM allows the application of image-like operations that are useful for highlighting different aspects of the surface. The most common are filters that perform terrain smoothing and gradient operations to visualize steepness and orientation, as shown in Fig. 9.42a–c. Fig. 9.42d depicts a so-called hillshade raster that improves visualization of the terrain by a chosen source (in position and angle) of illumination.

9.5.2 Orthophotos

Orthogonal Perspective

Orthophoto is a technical term reserved for an image that shows objects on a reference surface using an orthogonal perspective. The reference surface is a DEM that consists of points with three coordinates each (x, y, z) and that defines the Earth's surface. The orthogonal perspective means a vertical view of the ground above each pixel, which is typically used for maps. However, any image taken by a camera does not have this perspective, because the whole scene is photographed from one point, thus leading to a central perspective. Frame cameras have a central perspective in x and y (Fig. 9.43a). Line cameras operating on satellite platforms, and in some cases also on aerial platforms, produce images that have a central perspective along the line photographed at one moment but an orthogonal perspective along the flight track (Fig. 9.43b).

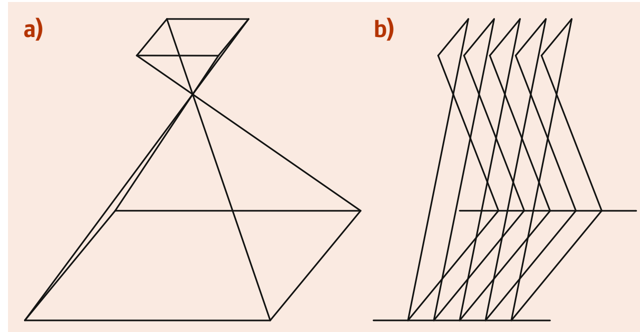


Fig. 9.43 Central perspective: (**a**) frame camera with central perspective in both dimensions, (**b**) line camera with central perspective across track and orthogonal perspective along track

An aerial image is distorted primarily for two reasons:

- The camera cannot be kept exactly horizontal when the photo is taken. Therefore, the roll and pitch components of attitude are not exactly zero, and, consequently, the image suffers from perspective distortion.
- Only in some exceptional cases is the Earth's surface flat. The central perspective of a camera causes height parallaxes that displace objects on higher ground towards the image border and objects on lower ground towards the image center.

Both effects are eliminated during orthophoto computation. In addition, the pixel size is set to a defined ground sample distance (GSD). This simplifies joint processing with vector data in later applications.

Other image errors, such as lens distortion, atmospheric refraction, and Earth curvature, are not eliminated during orthophoto computation, as their influence is usually considered in a basic part of the photogrammetric image-processing process that includes sensor calibration. Despite this, the geometry of the camera plays a prominent role in the computation of an orthophoto.

Rectification Methods

The orthographic projection is obtained from the central perspective through an analytical process called rectification.

9

Table 9.14 Overview of the rectification methods

Rectification method	Elevation model	Application
Perspective transformation	No model	Analytical plotter (approximation)
Polynomials		Satellite imagery
Standard orthophoto	DTM	Airborne imagery
True orthophoto	DTM + buildings	Airborne imagery (special application)

The complexity of this operation depends on the scale of imagery and the required degree of exactness. An overview of the different approaches is provided in Table 9.14.

The perspective transformation provides only an approximate solution to rectification and has been employed in the past when airborne photogrammetry made use of analog rectification instruments. The polynomials provide a somewhat better two-dimensional relation between the image and the ground. Such an approximation is usually sufficient for the rectification of satellite images. The transformation coefficients are commonly obtained through the identification of ground control points that are distributed across the image. This way, the images are simultaneously oriented and rectified.

By far most orthophotos are produced by making use of DEM (Sect. 9.5.1), especially DTMs that exclude vegetation and manmade structures. Consequently, the imaged buildings are rectified only at the base and not above the ground. Apart from the existence of the DTM, the prerequisite for correct rectification is the knowledge of image orientation parameters (interior and exterior) that shall be transformed to the same datum as the DTM employed. A common procedure for orthophoto creation using DEM is the following:

1. An empty orthophoto is created at the start. Such an orthophoto can be regarded as a grid or matrix with cells of predefined (pixel) size. Hence, knowing the coordinates (x and y) of the image-corner point, the geographical position of each pixel is uniquely defined by its row and column.
2. The geographical height (z-coordinate) of each pixel is identified through DEM as a function of its x- and y-position (e.g., by interpolation).
3. A vector is formed between the image origin and the x-y-z-coordinates of an empty pixel in the orthophoto (Fig. 9.44). This vector is intersected with the image through the collinearity condition to define its x'- and y'-photo-coordinates.
4. As the resulting x'- and y'-photo-coordinates do not necessarily correspond to the center of a pixel, its RGB-color (or gray) value on the orthophoto is found via interpolation with the neighboring pixels on the photograph. The interpolation can also be performed across several images, and these values can be further averaged to stabilize the resulting orthophoto radiometrically.

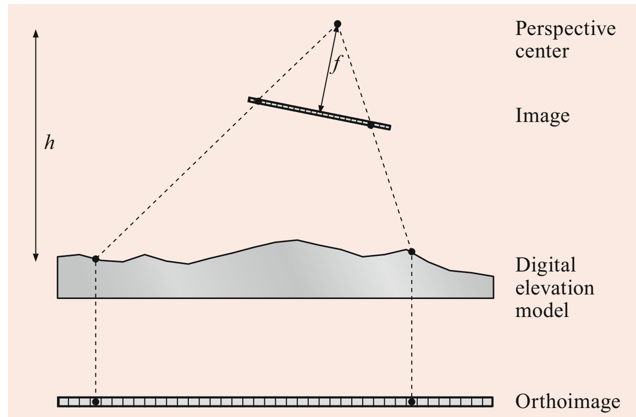


Fig. 9.44 The creation of an orthoimage with the use of DEM

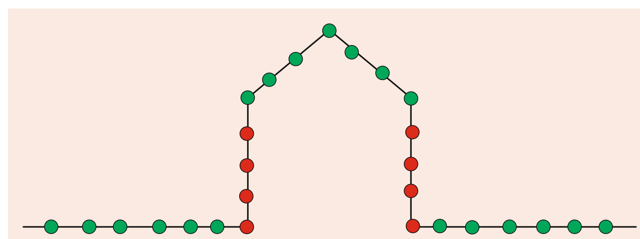


Fig. 9.45 Creation of a true orthophoto based on a dense point cloud (3-D). *Green points*: Points of the true orthophoto; *red points*: Points not used for the true orthophoto because they are not highest at their 2-D-position

Today, a *true orthophoto* is computed based on a dense point-cloud that represents a surface-model, i.e., mainly ground, buildings, and vegetation. Out of that point-cloud, all those points are deleted which are not the highest at any given position. The remaining points form the *true orthophoto* (Fig. 9.45).

References

1. Wolf, P.: Elements of Photogrammetry. McGraw-Hill, New York (1974)
2. Kraus, K.: Photogrammetry – Geometry from Images and Laser Scans. Walter de Gruyter, Berlin, Boston (2007)
3. Dorstel, C., Jacobsen, K., Stallmann, D.: DMC – Photogrammetric accuracy – Calibration aspects and generation of synthetic DMC images. In: Proc. Opt. 3-D Meas. Tech. Zurich, vol. VI, pp. 74–82. (2003)

4. Ladstadter, R., Tschemmernegg, H., Gruber, M.: Calibrating the ultracam aerial camera systems, an update. In: Proc. Int. Calibration Orientat. Workshop, EuroCOW, p. 8. (2010)
5. Gonzalez, R., Woods, R.: Digital Image Processing. Addison-Wesley, Reading (1992)
6. Ehlers, M., Klonus, S., Astrand, P., Rosso, P.: Multi-sensor image fusion for pansharpening in remote sensing. Int. J. Image Data Fusion **1**, 25–45 (2010)
7. Petri, G.: Deja vu – The configurations of the new airborne digital imagers are all rooted in the distant past! GeoInformatics **3**(5), 48–51 (2000)
8. Petri, G., Walker, S.: Airborne digital imaging technology: A new overview. Photogramm. Rec. **22**, 203–225 (2007)
9. Wehr, A., Lohr, U.: Airborne laser scanning an introduction and overview. ISPRS J. Photogramm. Remote Sens. **54**(2/3), 68–82 (1999)
10. Hug, C.: The scanning laser altitude and reflectance sensor – An instrument for efficient 3D terrain survey. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 30, pp. 100–107. (1994)
11. Lohr, U., Beraldin, A., Blais, F.: Laser scanning technology. In: Airborne and Terrestrial Laser Scanning. Whittles Publishing, Dunbeath (2010)
12. Schaer, P.: In-flight Quality Assessment and Data Processing for Airborne Laser Scanning. PhD thesis. The Swiss Federal Institute of Technology, Lausanne (2009)
13. Carabajal, C., Harding, D., Luthcke, S.B., Fong, W., Rowton, S., Frawley, J.: Processing of shuttle laser altimeter range and return pulse energy data in support of SLA-02. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 32-A4, pp. 269–277. (1999)
14. Spikes, V., Csatho, B., Whillans, I.: Airborne laser profiling of Antartic ice stream for change detection. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 32-3-W14, p. 7. (1999)
15. Geiger, A., Kahle, H.G., Limpach, P.: Airborne Laser Profiling. ETH Research Database 6120. Swiss Federal Institute of Technology, ETH Zurich, Zurich (2009)
16. Morin, K.: Calibration of Airborne Laser Scanners. M.Sc.. Univ. Calgary, Calgary (2002). UCGE Report No. 20179
17. Glennie, C., Lichti, D.: Static calibration and analysis of the velodyne HDL-64E S2 for high accuracy mobile scanning. Remote Sens. **2**, 1610–1624 (2010)
18. Glennie, C., Kusari, A., Facchin, A.: Calibration and stability analysis of the VLP-16 laser scanner. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XL-3/W4, pp. 55–60. (2016) https://doi.org/10.5194/isprsarchives-XL-3-W4-55-2016
19. Betz, J.: Engineering Satellite-Based Navigation and Timing. Wiley, IEEE Press, Hoboken, New Jersey (2016)
20. Satirapod, C., Homniam, P.: GPS precise point positioning software for ground control point establishment in remote sensing applications. J. Surv. Eng. **132**(1), 11–14 (2006)
21. Titterton, D., Weston, J.: Strapdown Inertial Navigation Technology (1997). Part of IEE radar, sonar, navigation and avionics series, Stevenage, U.K., all the details in strapdown inertial navigation, attention: some mistakes in equastions
22. Jekeli, C.: Inertial Navigation Systems with Geodetic Applications. Walter de Gruyter, Berlin (2001)
23. Andersson, R., Bilger, H., Stedman, G.: Sagnac effect: A century of Earth-rotated interferometers. Am. J. Phys. **62**, 975–985 (1994)
24. Khaghani, M., Skaloud, J.: Application of vehicle dynamic modeling in UAVs for precise determination of exterior orientation. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLI-B3, pp. 827–831. (2016) https://doi.org/10.5194/isprsarchives-XLI-B3-827-2016
25. Khaghani, M., Skaloud, J.: Assesment of VDM-based autonomous navigation of a UAV under operational conditions. Robot. Auton. Syst. **106**, 152–154 (2018). minor review pending
26. Weiss, J., Kee, D.: A direct performance comparison between loosely coupled and tightly coupled GPS/INS integration techniques. In: Proc. 51st Annu. Meet. Inst. Navig., Colorado Springs, pp. 537–544. (1995)
27. Wei, M., Schwarz, K.P.: Testing a decentralized filter for GPS/INS integration. In: Proc. IEEE Symp. Position Locat. Navig. (1990)
28. Scherzinger, B.: Precise robust positioning with inertially aided RTK. J. Inst. Navig. **53**(2), 73–83 (2006)
29. Greenspan, R.: Inertial navigation technology from 1970 to 1995. J. Inst. Navig. **42**, 165–186 (1995)
30. Legat, K.: Approximate direct georeferencing in national coordinates. ISPRS J. Photogramm. Remote Sens. **60**, 239–255 (2006)
31. Skaloud, J., Legat, K.: Theory and reality of direct georeferencing in national coordinates. ISPRS J. Photogramm. Remote Sens. **63**, 272–282 (2008)
32. Rehak, M., Skaloud, J.: Fixed-wing micro aerial vehicle for accurate corridor mapping. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. II-1/W4, pp. 23–31. (2015) https://doi.org/10.5194/isprsannals-II-1-W1-23-2015
33. Cramer, M., Stallmann, D.: System calibration for direct georeferencing. In: ISPRS Comm. III Symp. Photogramm. Comput. Vis., Graz, p. 6. (2002)
34. Kruck, E.: Combined IMU and sensor calibration with BINGO-F. In: Proc. OEEPE Workshop Integr. Sens. Orientat. pp. 84–108. (2001). CD-ROM
35. Skaloud, J., Schaer, P.: Towards a more rigorous boresight calibration. In: ISPRS Int. Workshop Theory Technol. Real. Inert. GPS Sens. Orientat., Castelldefels (2003)
36. Cosandier, D.: Generating a Digital Elevation Model and Orthomosaic from Pushbroom Imagery. Univ. Calgary, Calgary (1999). UCGE Report No. 20133
37. Tempelmann, U., Hinsken, L.: Triangulation of ADS40 pushbroom image blocks – Not much different from classical frame blocks? In: SGPBF Annu. Meet. (2005)
38. Burman, H.: Calibration and Orientation of Airborne Image and Laser Scanner Data Using GPS and INS. PhD thesis. Royal Institute of Technology, Stockholm (2000)
39. Kager, H.: Discrepancies between overlapping laser scanning strips – Simultaneous fitting of aerial laser scanner strips. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 35, pp. 555–560. (2004)
40. Morin, K., El-Sheimy, N.: Post-mission adjustment of airborne laser scanning data. In: FIG XXII Int. Congr. Washington, p. 12. (2002)
41. Friess, P.: Toward a rigorous methodology for airborne laser mapping. In: Proc. EuroCOW, p. 7. (2006). on CDROM
42. Skaloud, J., Lichti, D.: Rigorous approach to bore-sight self calibration in airborne laser scanning. ISPRS J. Photogramm. Remote Sens. **61**, 47–59 (2006)
43. Longuet-Higgins, H.C.: A computer algorithm for reconstructing a scene from two projections. Nature **293**, 133–135 (1981)
44. Hartley, R.I.: In defence of the 8-point algorithm. IEEE Trans. Pattern Anal. **19**(6), 580–593 (2012)
45. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. Int. J. Comput. Vis. **37**(2), 151–172 (2000)
46. Jazayeri, I., Fraser, C.: Interest operators for feature-based matching in close range photogrammetry. Photogramm. Rec. **25**(129), 24–41 (2010)

47. Haralick, R.M., Shapiro, L.G.: Computer and Robot Visions. Addison-Wesley, Reading (1992)
48. Förstner, W., Gülch, E.: A fast operator for detection and precise location of distinct points, corners and centers of circular features. In: Proc. ISPRS Intercomm. Workshop Fast Process. Photogramm. Data, pp. 281–305. (1987)
49. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. 4th Alvey Vis. Conf. pp. 147–151. (1988)
50. Lowe, D.G.: Distinctive image features from scale invariant key points. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
51. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1615–1630 (2005)
52. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications ACM **24**, 381–395 (1981)
53. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. Pattern Anal. Mach. Intell. **26**(6), 756–770 (2004)
54. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer, London (2010)
55. Hirschmüller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. **30**(2), 328–341 (2008)
56. Haala, N.: Dense Image Matching Final Report, pp. 115–145 (2014). Tech. Rep., EuroSDR Official Publication No. 64
57. Colomina, I., Blazquez, M.: A unified approach to static and dynamic modeling in photogrammetry and remote sensing. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 35-B1, pp. 178–183. (2004)
58. Kerstling, A.P., Habib, A., Bang, K.I., and Skaloud, J.: Automated approach for rigorous light detection and ranging system calibration without preprocessing and strict terrain coverage requirements. Opt. Eng. **51**(7), 076201 (2012)
59. Strasdat, H., Montiel, J., Davison, A.: Real-time monocular slam: Why filter? In: IEEE Int. Conf. Robot. Autom. (ICRA), pp. 2657–2664. (2010)
60. Cucci, D.A., Rehak, M., Skaloud, J.: Bundle adjustment with raw inertial observations in UAV applications. ISPRS J. Photogramm. Eng. Remote Sens. **130**, 1–12 (2017)
61. Freyer, J., Brown, D.: Lens distortion for close-range photogrammetry. Photogramm. Eng. Remote Sens. **52**, 51–58 (1986)
62. Gruen, A.: The accuracy potential of the modern bundle block adjustment in aerial photogrammetry. Photogramm. Eng. Remote Sens. **48**, 45–54 (1982)
63. Rehak, M., Skaloud, J.: Time synchronization of consumer cameras on micro aerial vehicles. ISPRS J. Photogramm. Remote Sens. **123**(1), 114–123 (2017)
64. Förstner, W., Wrobel, B.: Mathematical concepts in photogrammetry. In: McGlone, J., Mikhail, E. (eds.) Manual of Photogrammetry, 5th edn., pp. 15–180. ASPRS, Bethesda (2004)
65. Bjerhammar, A.: Theory of Errors and Generalized Matrix Inverses. Elsevier, Amsterdam (1973)
66. Kummerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: g^2o: A general framework for graph optimization. In: IEEE Int. Conf. Robot. Autom. (ICRA), pp. 3607–3613. IEEE, New York (2011)
67. Davis, T.A.: Direct Methods for Sparse Linear Systems. SIAM, Philadelphia (2006)
68. Cucci, D., Skaloud, J.: On inertial measurements in dynamics networks. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. IV-2/W5, pp. 549–557. (2019)
69. Hoffman, O., Nave, P., Ebner, H.: DPS – A digital photogrammetric system for producing digital elevation models and orthophotos by means of linear array scanner imagery. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 24, pp. 216–227. (1982)
70. Cramer, M.: The ADS40 Vaihingen/Enz geometric performance test. ISPRS J. Photogramm. Remote Sens. **60**, 363–374 (2006)
71. Ebner, H.: Self calibrating block adjustment. In: XIIIth Congr. Int. Soc. Photogramm. Com. V., Stockholm (1978)
72. Filin, S., Vosselman, G.: Adjustment of airborne laser altimetry strips. In: ISPRS XXth Congr. Geo-Imagery Bridg. Cont. pp. 285–289. (2004)
73. Glira, F., Pfeifer, N., Mandelburger, G.: Hybrid orientation of airborne lidar point clouds and aerial images. In: he International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. IV-2/W5, pp. 567–574. (2019)
74. Filin, S.: Recovery of systematic biases in laser altimetry data using natural surfaces. Photogramm. Eng. Remote Sens. **69**(11), 1235–1242 (2003)
75. El-Sheimy, N., Valeo, C., Habib, A.: Digital Terrain Modeling. Artech House, Norwood (2005)

Jan Skaloud Jan Skaloud is a senior scientist and lecturer at the Swiss Federal Institute of Technology Lausanne (EPFL) and Adjunct Professor in the Department of Geomatics Engineering at The University of Calgary. His research is related to satellite and inertial navigation, sensor calibration, and modern mapping techniques with active and passive optical sensors. He is a member and active participant in numerous international scientific organizations in this field.

Michael Cramer Michael Cramer is Senior Lecturer at the Institute for Photogrammetry (ifp), University of Stuttgart, Germany. He received his Doctor's degree in Geodesy and Geoinformatics from the same university. His research was previously devoted to georeferencing of sensors using integrated GPS/inertial data for airborne photogrammetry. Currently, his focus is on digital photogrammetric airborne sensor systems with special emphasis on their overall accuracy. He also is active in standardization activities worldwide.

Norbert Haala Norbert Haala is Associate Professor at the Institute for Photogrammetry, the University of Stuttgart, where he is responsible for lectures in the field of photogrammetric image processing. His research interests include virtual city models and image-based 3-D reconstruction.

Remote Sensing

10

Erik Borg, Sina C. Truckenbrodt, Angela Lausch, Peter Dietrich, and Karsten Schmidt

Contents

E. Borg (✉)
German Remote Sensing Data Center, German Aerospace Center (DLR)
Neustrelitz, Germany
e-mail: erik.borg@dlr.de

S. C. Truckenbrodt
Institute of Geography, Earth Observation, Friedrich Schiller University Jena
Jena, Germany
e-mail: sina.truckenbrodt@uni-jena.de

A. Lausch
Department of Computational Landscape Ecology, Helmholtz Centre for Environmental Research – UFZ
Leipzig, Germany
e-mail: angela.lausch@ufz.de

P. Dietrich
Monitoring and Exploration Technologies, Helmholtz Centre for Environmental Research - UFZ
Leipzig, Germany
e-mail: peter.dietrich@ufz.de

K. Schmidt
Remote Sensing Technology Institute, German Aerospace Center (DLR)
Neustrelitz, Germany
e-mail: Karsten.Schmidt@dlr.de

Abstract

Remote sensing is an essential method of collecting the geoinformation needed for investigations of the status or dynamics of environmental and social systems and processes. In this context, the multifunctional concept underlying remote sensing plays an important role. In particular, the removal of obstacles such as inadequate data availability or excessively long processing times when gathering information established remote sensing as a stable data source within the framework of environmental monitoring. The term *remote sensing* covers a multitude of different sensors, measuring principles, recording modes, and carrier platforms as well as combinations thereof. This makes it possible to generate data with various geometric, spectral, radiometric, and temporal resolutions to solve different problems. In the following, the physical basics of remote sensing (such as the theoret-

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_10

ical basics and laws of electromagnetic radiation and the interaction of that radiation with natural materials), the technical basics of sensors, the properties of data (including aspects of data processing, processors, and processing chains as well as information extraction from remote sensing data), and selected areas of application of the results from remote sensing are explained.

Keywords

remote sensing · physical basics · technical basics · sensors · data · data acquisition · data correction · data processing

10.1 What Is Remote Sensing?

Remote sensing (RS) is the contactless observation and measurement of the energetic and polarimetric characteristics of inherent and reflected radiation from elements and objects that constitute land surfaces, oceans, and atmospheres. In various electromagnetic wavelength ranges, RS can help to describe locations, the nature and temporal variability of natural parameters, phenomena, resources, the environment, and anthropogenic objects and formations [1, 2]. In that sense, RS includes the "entirety of methods/procedures for obtaining information about the Earth's surface or other objects that are not directly accessible" (DIN 18716), including phenomena on the Earth's surface or in the atmosphere [3]. The deployed methods/procedures utilize sensors that are not in physical contact with the observed object/phenomenon to perform nondestructive measurements with the help of electromagnetic signals, which are then evaluated qualitatively and quantitatively.

RS data are often the basis for operative decisions and action recommendations and must therefore be provided quickly and in a user-friendly format. Also, the personnel and financial effort needed to process the data should be acceptable. In this context, it is of primary importance to collect data about the spatial and temporal distributions of physical parameters or phenomena that are suitable for examining the structural and material composition of the object and/or phenomenon of interest.

The multifunctional concept of RS is of particular importance in environmental applications, and includes the following aspects [1]:

- *Multispectral exploration*: Observing in different wavelength ranges to identify the spectral characteristics of objects,
- *Multispatial exploration*: Investigating an object and/or a phenomenon using different spatial resolutions,
- *Multistationary exploration*: Observing objects/phenomena from different positions (e.g., different orbit heights and viewing angles),
- *Multipolarized exploration*: Observing and identifying the polarization properties of objects,
- *Multimission exploration*: Investigating an object and/or a phenomenon via multiple RS missions that utilize the same and/or different sensors,
- *Multitemporal exploration*: Observing at different times to capture the dynamic behavior of objects and/or phenomena,
- *Multidisciplinary*: Using the same data stock for different applications in the environmental sciences and geosciences (ecology, geological prospecting, security policy, environmental policy, etc.).

Considering this multifunctional concept, RS can make significant contributions to (i) the identification, characterization, and documentation of environmental conditions (status, damage, etc.), (ii) basic research to facilitate the quantitative and qualitative modeling of cause–effect complexes, (iii) the conception, verification, and control of countermeasures, and (iv) a new understanding of the environment and the subjective perception of risk that is appropriate for the environmental situation of interest.

Depending on the properties of the electromagnetic waves and their interactions with objects and phenomena of interest, the following parameters can be used to derive information: intensity, runtime (distance between the sensor and the object or phenomenon), wavelength and spectral information (provided that the spectral characteristics of the object or phenomenon are known), polarization, phase, spatial information at the macroscopic level (objects can be considerably larger than a wavelength), temporal information (by recurrently observing the object or phenomenon of interest from the same position, etc.), and angular information [4].

Sensors for RS can be installed on lifting platforms, towers, balloons, kites, unmanned aerial vehicles (UAVs), helicopters, aircrafts, and satellites space stations, and can have different spatial, temporal, radiometric, and spectral resolutions. The optimization of sensor and platform parameters is task oriented.

10.2 Theoretical Background

Electromagnetic waves are transverse waves and do not require a transmission medium to propagate. Indeed, they reach their maximum propagation speed in a vacuum (the vacuum speed of light $c_0 = 299\,792\,458\,\mathrm{m/s}$). When propagating through matter, their velocity reduces to c_m, where $c_\mathrm{m} < c_0$. Two characteristic parameters of electromagnetic waves are the wavelength λ and the frequency f. The wavelength of a periodic wave is the distance traveled by the wave before the wave pattern reoccurs. The wavelength is expressed in meters or a multiple thereof. The frequency is inversely proportional to the wavelength ($f = c/\lambda$, $c = (c_0, c_m)$) and describes the number of times that the periodic pattern occurs per time unit. It is expressed in Hz or a multiple of this.

10.2.1 The Nature of Electromagnetic Waves

Electromagnetic waves are characterized by the following properties

- They are always accompanied by an electric field and a magnetic field that are orthogonal to each other and oscillate perpendicularly to the direction of wave propagation,
- They can be reflected and refracted (i.e., scattered),
- They can generate interference and diffraction,
- They can be described by their wavelength/frequency, amplitude, phase, propagation direction, velocity, polarization, and coherence.

10.2.2 Radiation Laws

Any body with a temperature T above absolute zero ($=$ 0 K or $-273.15\,°\mathrm{C}$) emits radiation in the form of electromagnetic waves. The physical basics that are relevant for capturing spectral signatures by RS are described by, for example, the radiation laws [1, 5]. These laws are based on the concept of an idealized black body that absorbs all incident radiation regardless of its wavelength and intensity. This means that such a body does not reflect radiation. The spectrum emitted by a black body (with $T > 0\,\mathrm{K}$) depends on its temperature and is described by *Planck's law of radiation.*

Consider the radiation energy $\mathrm{d}E(\lambda, T)$ emitted by an arbitrary body within the wavelength range $[\lambda, \lambda + \mathrm{d}\lambda]$ per time unit t and surface area A. The emissivity $\varepsilon(\lambda, T)$ (in $\mathrm{W\,m^{-3}}$) of the body can be defined by

$$\varepsilon(\lambda, T) = \frac{\mathrm{d}E(\lambda, T)}{A \cdot t\,\mathrm{d}\lambda} . \tag{10.1}$$

On the other hand, the absorptivity α is defined by the amount $\mathrm{d}E_\mathrm{A}$ of the incident electromagnetic energy $\mathrm{d}E_0$ that is absorbed by the body within the wavelength range $[\lambda, \lambda + \mathrm{d}\lambda]$ per time unit t and surface area A via

$$\alpha(\lambda, T) = \frac{\mathrm{d}E_\mathrm{A}}{\mathrm{d}E_0} . \tag{10.2}$$

This is a dimensionless number. For a black body, it holds that $\alpha = 1$.

Kirchhoff's radiation law states that, for each emitter in thermodynamic equilibrium, the ratio of the emissivity ε to the absorptivity α at a given wavelength λ and temperature T is constant and equal to the emissivity $\varepsilon_\mathrm{blackbody}$ of a black body at the same temperature and wavelength, i.e.,

$$\varepsilon(\lambda, T)/\alpha(\lambda, T) = \varepsilon_\mathrm{blackbody}(\lambda, T) . \tag{10.3}$$

The emissivity $\varepsilon_\mathrm{blackbody}$ of a black body is described by *Planck's law of radiation*,

$$\varepsilon_\mathrm{blackbody}(\lambda, T) = \frac{2\pi h {c_0}^2}{\lambda^5}\left(\frac{1}{\mathrm{e}^{hc_0/(\lambda k_\mathrm{B} T)} - 1}\right) . \tag{10.4}$$

Here, h denotes the Planck constant

$$h = 6.626070040 \times 10^{-34}\,\mathrm{J\,s}$$

and k_B is the Boltzmann constant [6]

$$k_\mathrm{B} = 1.38064852 \times 10^{-23}\,\mathrm{J/K} .$$

Planck's law of radiation describes the spectral distribution of the electromagnetic energy $\varepsilon_\mathrm{blackbody}(\lambda, T)$ (in $\mathrm{W\,m^{-3}}$) emitted in the form of radiation by a black body in thermodynamic equilibrium.

Sometimes the ratio of the emissivities of an arbitrary body and a black body, i.e.,

$$\varepsilon'(\lambda, T) = \frac{\varepsilon(\lambda, T)}{\varepsilon_\mathrm{blackbody}(\lambda, T)} , \tag{10.5}$$

is considered. Using this definition, *Kirchhoff's radiation law* takes the simple form

$$\varepsilon'(\lambda, T)/\alpha(\lambda, T) = 1 . \tag{10.6}$$

It is seen from *Planck's law* that there are radiation maxima, and the positions of these maxima depend on the temperature. Determining the corresponding maximum wavelength λ_max by solving an extreme value problem yields *Wien's displacement law*, which states that the wavelength of the spectral peak λ_max is inversely proportional to the temperature T (in K) of a black body, in other words

$$\lambda_\mathrm{max} = k_\mathrm{W} T^{-1} . \tag{10.7}$$

Here, k_W, known as Wien's displacement constant, is $2.8977729 \times 10^{-3}\,\mathrm{m\,K}$.

The total energy emitted by a black body in the form of radiation at temperature T is obtained by integrating *Planck's law* of radiation over the whole spectrum. The result is termed the *Stefan–Boltzmann law*, which states that the total radiation energy E emitted by a black body per time unit

t and surface area A (which equals the flux density or radiant exitance M) is proportional to the fourth power of its temperature T, i.e.,

$$M = \frac{E}{At} = \sigma T^4 \,. \tag{10.8}$$

The scaling constant is known as the Stefan–Boltzmann constant $\sigma = 5.6697 \times 10^{-8}\,\mathrm{Wm^{-2}\,K^{-4}}$.

A real or *gray body* emits less energy than a black body. A gray body is defined as having the same absorptivity at all wavelengths. For such a body, the *Stefan–Boltzmann law* reads

$$M = \varepsilon' \sigma T^4 \,. \tag{10.9}$$

10.2.3 Electromagnetic Spectrum and Remote Sensing Working Ranges

Contactless observations can be performed, for example, by airborne (e.g., an aircraft, helicopter, UAV, or balloon) or spaceborne (i.e., a satellite) RS sensors of passive (e.g., a camera, scanner, radiometer) or active (e.g., light detection and ranging (lidar) and radio detection and ranging (radar)) systems. The sensors record the radiation that is reflected, scattered, or emitted from the land surface, the ocean, or in the atmosphere.

When applied to the Earth (i.e., Earth observation), RS is based on the detection of radiant energy that (i) is emitted by the Earth itself (black body with $T_{\text{surface Earth}} = 300\,\mathrm{K}$ [3] or $T_{\text{fire}} = 600$–$1000\,\mathrm{K}$ [7]; see Fig. 10.1a) or is reflected and scattered at the Earth's surface, (ii) originates from the Sun (black body with $T_{\text{surface Sun}} = 5777\,\mathrm{K}$; Fig. 10.1a [8]), or (iii) is a transmitter illuminating the Earth artificially.

The radiant energy received by the sensor depends on, for example, the temperature of the Sun and the Earth (see Fig. 10.1a for the case that the Sun and Earth are approximated as black bodies) and the permeability of the atmosphere (Fig. 10.1b, c) to different parts of the electromagnetic spectrum (Fig. 10.1d; Table 10.1).

The permeable regions of the electromagnetic spectrum—those in which the atmosphere is partially transparent to electromagnetic waves (Fig. 10.1c)—are called *atmospheric windows* [3]. These regions allow phenomena caused by the interaction of radiation with atoms, molecules, structures, etc., on the Earth to be recorded or measured from remote distances through the atmosphere (Fig. 10.1e). All of these regions are currently exploited by diverse RS systems (Fig. 10.1f). These regions appear in the spectrum due to the reflection, transmission, absorption, and scattering properties of the particles and molecules that constitute the atmosphere (especially H_2O, CO, CO_2, O_3 and dust; see Fig. 10.1b [6, 15]), as they determine the spectral characteristics of the atmosphere. While shortwave radiation ($\lambda < 300\,\mathrm{nm}$) is almost completely absorbed, parts of the visible (VIS), infrared (IR), and microwave (MW) regions are permeable enough to be penetrated by electromagnetic waves (Fig. 10.1c).

10.2.4 Interaction of Electromagnetic Radiation with Natural Media

When electromagnetic waves impinge upon the atmosphere and/or the Earth's surface, they are scattered by inhomogeneities and/or particles in the atmosphere, the soil, or water (volume scattering), or by boundaries such as soil or water surfaces (surface scattering). Additionally, some or all of the incident waves can be absorbed. Both scattering and absorption contribute to extinction.

The amount of scattering and (more generally) extinction that occurs depends strongly on the size of the scattering object or structure compared to the wavelength λ of the incoming radiation. In order to quantify this ratio between the size and the wavelength, the size parameter

$$k_0 a = (2\pi a)/\lambda \tag{10.10}$$

is introduced. Here, a denotes a typical size such as the radius of a spherical particle, the extent of the inhomogeneity, or a length parameter to describe the surface roughness.

$$k_0 = (2\pi)/\lambda \tag{10.11}$$

is the free-space wavenumber. The larger the size parameter, the stronger the scattering and extinction. Moreover, dedicated approaches can be developed and applied to describe the scattering behavior, depending on the value of the size parameter.

For example, consider the elastic scattering of a plane electromagnetic wave on a homogeneous and isotropic spherical particle. In contrast to inelastic scattering, elastic scattering means that the wavelengths of the incident and scattered fields are equal, i.e., the wavelength does not change during the scattering process. If the sphere is much smaller than the wavelength of the incident plane wave ($k_0 a \ll 1$), the *Rayleigh approximation* can be applied. The incident plane wave induces a dipole moment in the scattering particle so that the scattered field is the field of a dipole, characterized by dipole polarizability [16]. Thus, this scattering process is also referred to as *Rayleigh scattering*. It should be noted that this approach can be applied to not only homogeneous spheres but also inhomogeneous nonspherical particles, provided that the size parameter of the latter is much smaller than 1. The actual shapes and compositions

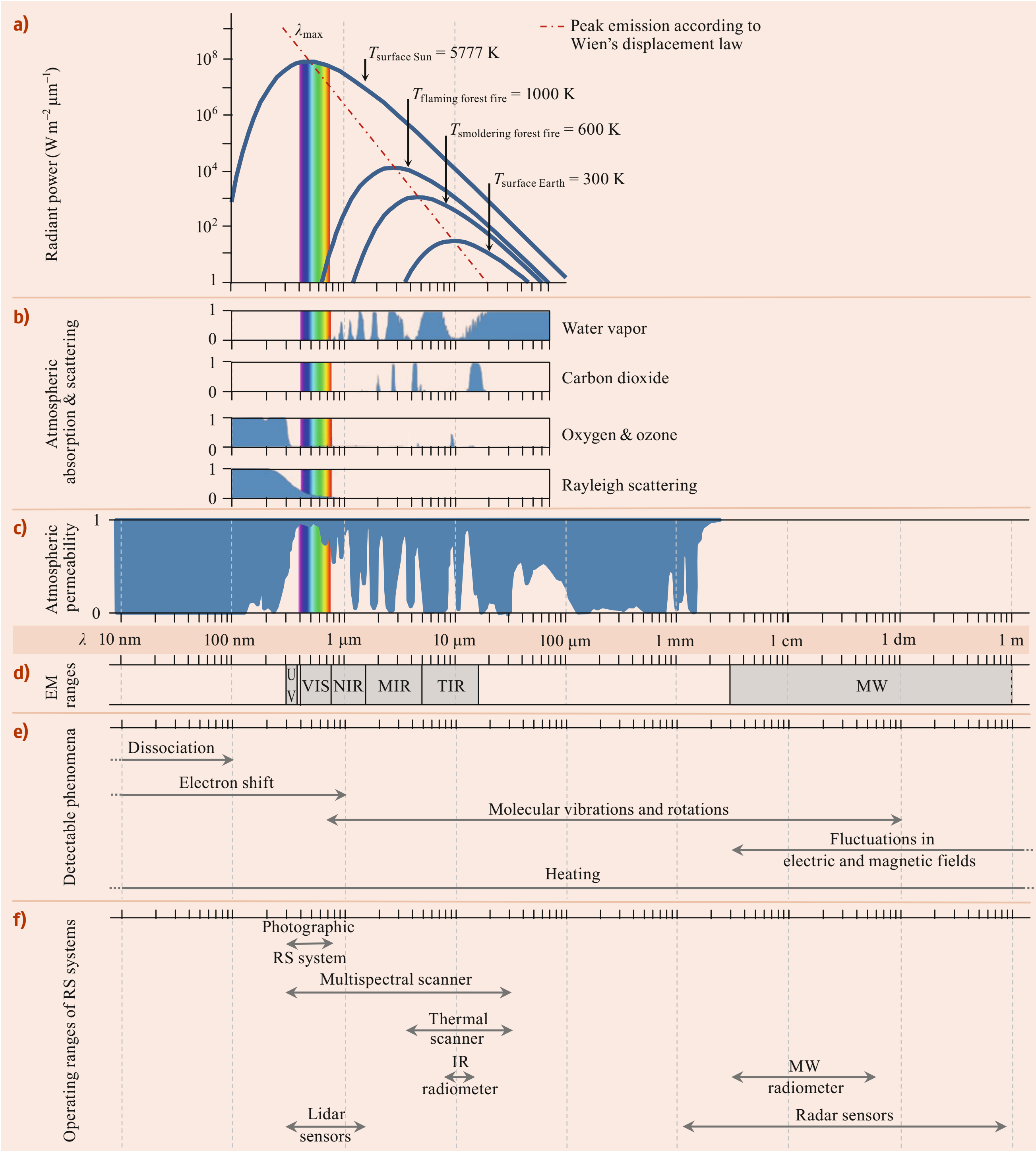


Fig. 10.1 From emitter to receiver. (**a**) Radiation sources are approximated by a black body to describe the distribution of the radiant power (10.4) they emit over the electromagnetic (EM) spectrum [7, 8]. (**b**) Molecules and particles of the atmosphere hamper the propagation of radiation through the Earth's atmosphere at certain wavelengths [6, 9]. (**c**) The ranges of the spectrum in which the atmosphere is permeable to radiation [3] are called atmospheric windows. They correspond to various (**d**) spectral regions [10–12] and allow for (**e**) the detection of various radiation-induced phenomena [13] by (**f**) diverse types of remote sensing (RS) systems [3, 5, 14]

Table 10.1 Main ranges of the electromagnetic spectrum in which RS is performed (adapted from [10–12])

Range	Abbreviation	Wavelength range	Details	
Ultraviolet	UV	0.10–0.40 µm	Very narrow zone of electromagnetic radiation	
Visible	VIS	0.40–0.75 µm	Violet	0.400–0.446 µm
			Blue	0.446–0.500 µm
			Green	0.500–0.578 µm
			Yellow	0.578–0.592 µm
			Orange	0.592–0.620 µm
			Red	0.620–0.750 µm
Near-infrared	NIR	0.75–1.50 µm	Frequently used in RS	
Mid-infrared	MIR	1.50–5.0 µm	SWIR (shortwave infrared)	1.5–3.0 µm
Thermal infrared	TIR	5.0–15.0 µm	Most of the energy of this long-wavelength radiation is emitted from the Earth	
Microwave	MW	0.30–100 cm	P	76.90–100.00 cm
			L	19.35–76.90 cm
			S	7.14–19.35 cm
			C	5.22–7.14 cm
			X	2.75–5.22 cm
			K	0.83–2.75 cm
			Q	0.65–0.83 cm
			V	0.54–0.65 cm
			W	0.30–0.54 cm
			This is the radiation with the longest wavelengths utilized in both active and passive RS. MW signals can penetrate clouds and fog	

of the particles can be taken into account via appropriately chosen dipole polarizabilities.

If the radius of the spherical particle is comparable to or larger than the wavelength ($k_0 a \geq 1$), the incident plane wave induces multipole moments in the particle. A rigorous solution of this scattering problem was first provided by Gustav Mie in 1908, based on an analytical solution of *Maxwell's equations* for scattered and internal fields (*the Mie theory*; [16]). Therefore, this interaction process is also referred to as *Mie scattering*. In the limiting case of very small spheres, the *Rayleigh approximation* follows. Again, we should emphasize that the Mie theory only holds for plane electromagnetic wave scattering on homogeneous and isotropic spheres. However, the term *Mie scattering* is often used for any particle in this size parameter range, regardless of its actual shape and composition. The development of approaches to deal with nonspherical and inhomogeneous particles is an ongoing area of research [17].

If the radius of the spherical particle is much larger than the wavelength ($k_0 a \gg 1$) then the *geometric optics (GO) approximations* (also known as the ray-tracing or ray-optics approximations) can be used to model the scattering process. Here, the incident wave is represented by a collection of parallel wavelength-independent rays. These rays undergo a number of reflection, refraction, transmission, and absorption processes. On the particle surface, specular reflection occurs since the large spherical surface can be considered smooth. This means that the angle between the incident ray and the normal to the surface is equal to the angle between the reflected ray and the normal. The incident and reflected rays as well as the normal lie in the same plane. Furthermore, the refraction is described by *Snell's law*,

$$\frac{n_2}{n_1} = \frac{\sin \beta_1}{\sin \beta_2} \,. \tag{10.12}$$

The quantities n_1 and β_1 denote the refractive index and the incident angle, respectively, in medium 1. On the other hand, n_2 and β_2 are the refractive index and the angle of refraction, respectively, in medium 2. In order to obtain correct results using the *GO approximations*, the diffraction of the incident wave on the particle projection has to be added. As already mentioned in Sect. 10.2.1, this is a wave phenomenon through which a wavelength dependence enters the model (for more details, see [17]). It should be noted that the *GO approximations* can be also applied to describe the interaction of radiation with arbitrarily shaped and arbitrarily oriented large particles and large surface structures.

Figure 10.2 illustrates the ranges of applicability of the different approaches (the *Rayleigh approximation*, the *Mie theory*, and the *GO approximations*). It shows the extinction efficiency (i.e., the ratio of the extinction cross-section to the geometric cross-section) as a function of the size parameter for a nonabsorbing sphere with a real refractive index of 1.5. More generally, Fig. 10.2 also demonstrates that the interaction of electromagnetic radiation with a given particle depends strongly on the wavelength of the radiation.

It should be noted that the scattered field of any arbitrarily shaped three-dimensional (3-D) particle in free space forms

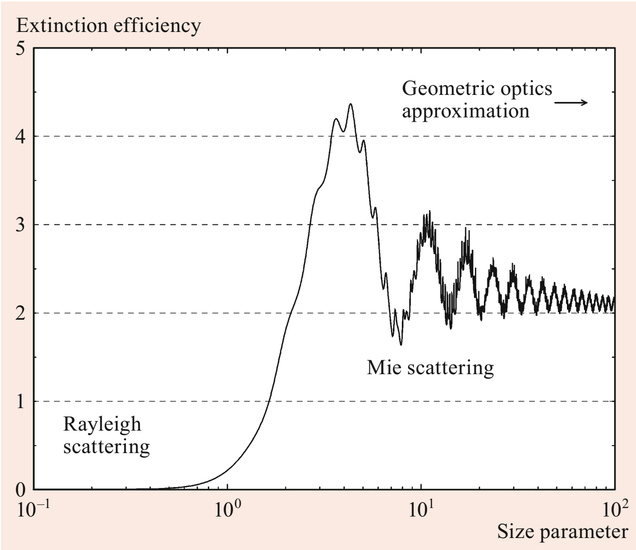


Fig. 10.2 Extinction efficiency of a homogeneous, isotropic, and non-absorbing sphere with a real refractive index of 1.5 as a function of the size parameter, as computed by implementing the Mie theory

an outgoing spherical wave far from this particle, i.e., in the far field zone of the scatterer. The field strength of the spherical wave decreases with $1/r$, where r is the distance to the scattering particle. Thus, the intensity of the spherical wave decreases with $1/r^2$. This can be generalized to the *inverse square law*. The intensity of the radiation uniformly emitted by a source into the 3-D free space is distributed on a spherical surface and decreases with $1/r^2$. A scattering particle can be regarded as one such source that is excited by the incident field to emit the scattered field.

The interaction of radiation with inhomogeneities and surfaces can be described by the same or similar methods as described above for the particles, depending on their dimensions (i.e., the size parameter). For instance, smooth, medium, and rough surfaces can be treated using *Maxwell's equations* or the *GO approximations*. If these surfaces also have small-scale roughness, the scattering problem can be treated within the framework of, e.g., perturbation approximations. The scattering at the underlying surface can be regarded as the unperturbed problem, while the scattering contribution of the small-scale roughness can be added as a first-order perturbation. In this way, the diffuse scattering or reflection that occurs in natural media can be modeled.

Note that various criteria are used to define the smoothness or roughness of a surface. For example, let h_A be the amplitude of the roughness and Θ_i the angle between the horizontal plane of the sensor carrier and the observed point in the terrain (depression angle). Then, a surface is considered to be smooth in the visible spectral range if the *Rayleigh criterion*

$$h_A \leq \lambda/(8\cos\Theta_i) \qquad (10.13)$$

is fulfilled [18]. In the microwave region, the *Fraunhofer criterion*

$$h_A \leq \lambda/(32\cos\Theta_i) \qquad (10.14)$$

was proposed by [19].

The amount of scattering and extinction depends not only on the size of the scattering object or structure (i.e., on the size parameter) but also on its dielectric properties, which are, in general, strongly wavelength dependent. The stronger the difference between the dielectric constant of an object and the dielectric constant of the surrounding medium at a given wavelength, the stronger the scattering and extinction. This can in turn influence the choice of the methods used to describe the interaction with the incoming radiation.

Consider, for instance, the illumination of terrain by radar waves. Radiation that hits a boundary between two dielectrics is partly reflected, refracted, and absorbed to a degree that depends on the electrical conductivities and relative dielectric constants of the media. Increasing the dielectric constant of the lower medium, i.e. that of terrain, causes a higher proportion of the radiation to be reflected. Approximate values of the electrical conductivity and the relative dielectric constants for different natural materials in the microwave region are listed in Table 10.2.

A comparison of the dielectric constants of dry and wet soils shows that the electrical behavior of soils is dominated by its moisture content. In the microwave range, the real part of the relative dielectric constant $\Re(\varepsilon)$ of dry soil minerals is $2 \leq \Re(\varepsilon) \leq 6$, while $\Re(\varepsilon)$ values of > 80 are reached for water [20, 21].

In nearly all RS scenarios, one is faced with a variety of different scattering objects and ensembles that are interacting with radiation, rather than a single type of particle, inhomogeneity, or surface. Different approaches have been developed to cope with this problem. From the experimental side, dedicated scattering experiments can be performed, and the results from them can be parameterized so that they are applicable to situations other than those in the actual setups. From the theoretical point of view, models can be developed to treat different multiple scattering problems, e.g., multiple scattering in particle systems or multiple scattering between particles and surfaces in their vicinity. A frequently used program is the Multiple Sphere T-Matrix (MSTM) implementation by [26]. This calculates the electromagnetic scattering and absorption properties of a system of spherical particles, which can also have spherical inclusions. However, the gap between such models and its applicability for RS is, in general, still too large due to idealised and restrictive model assumptions. Multiple scattering approaches are quite complex in the electromagnetic case, even for spherical particles. MSTM only accounts for elastic scattering processes—inelastic scattering processes that may also be present, such as Raman scattering, are not considered. The resonant absorption of electromag-

Table 10.2 Approximate values of the electrical conductivity and relative dielectric constant [20–23] and thermal properties (adapted from [24, 25]) for various natural materials

Natural medium		Electrical properties		Thermal properties			
		Electric conductivity	Relative dielectric constant	Density	Specific heat	Thermal conductivity	Thermal diffusivity
		σ	ε	ρ	c_H	k	κ
		(S/m)	(–)	(10^3 kg/m^3)	(kJ/(kg K))	(W/(m K))	($\times 10^{-6}$ m^2/s)
Soil components	Snow (firn)	10^{-5}–10^{-2}	4–8	0.55–0.83	2.09	2.10–2.76	0.1–0.8
	Permafrost	10^{-6}–10^{-5}	1.4	2.2	1.741–2.097	0.1–4.0	0.1–80
	Clay minerals	1.8–2.8		2.65	0.90	2.92	1.22
	Organic matter	2.8		1.30	1.92	0.25	0.10
	Pure freshwater	10^{-4}	81	1.00	4.18	0.57	0.14
	Air	0	1	0.0012	1.01	0.025	20.50
Soil, sediments (dry–moist)	Sandy soil	1.4×10^{-4}–6.9×10^{-3}	2.6–25	1.60–2.00	0.80–1.48	0.24–0.74	0.24–0.74
	Clayey soil	1.1×10^{-4}–2.1×10^{-2}	2.5–19	1.60–2.00	0.89–1.55	0.18–0.51	0.18–0.51
	Loamy soil	2.7×10^{-4}–5.0×10^{-2}	2.4–15				
	Peat soil	0.025–0.05	43–69	0.26–1.06	1.92–3.65	0.10–0.12	0.10–0.12
Rocks (dry–moist)	Sandstone	1.6×10^{-9}–1	4.7–12	2.5	0.796	1.55	1.3
	Limestone	10^{-9}–2.5×10^{-2}	7–8	2.5	0.712	2.01	1.1
	Granite	10^{-8}–10^{-3}	5–7	2.6	0.671	2.93	1.6
	Basalt	10^{-2}	8	2.8	0.838	2.09	0.9

netic radiation (e.g., by gas molecules in the atmosphere) and many other effects (such as those caused by temperature and pressure changes) are also not taken into account.

Dedicated radiative transfer (RT) models are more suitable for approximately describing the propagation of electromagnetic radiation through the atmosphere and at the Earth's surface, as well as its interactions with different media (e.g., [27]). By applying these approximations, the limitations mentioned above are no longer a concern. If necessary, elastic scattering on particles is incorporated via phase functions in scalar RT models and phase matrices in vector RT approaches which are based on the abovementioned single-scattering approaches. Scalar RT models describe the propagation of scalar intensities, whereas vector RT models deal with the propagation the Stokes vector to account for the polarization state of electromagnetic radiation. Scattering at different boundaries and surfaces can also be taken into account within RT models.

In active remote-sensing scenarios based on radar or lidar systems, the intensity of the emitted radar or laser (light amplification by stimulated emission of radiation) radiation is, in general, much larger than that of the solar radiation or the radiation emitted by media in the atmosphere and on the Earth. Therefore, special models—the radar and lidar equations, respectively—are used to describe the propagation of these signals and to derive information on the atmosphere and the Earth's surface from these measurements. Radar signals and laser light are scattered by various inhomogeneities, particles, and surfaces. In simulations, the corresponding scattering quantities such as extinction and backscattering cross-sections can again be computed using the abovementioned scattering approaches. In certain RS scenarios, the solar light or the radiation emitted by media are also detected by the sensors, degrading the signal-to-noise ratio of the measurements. In these cases, special correction measures must be performed.

Table 10.3 Optical behavior of various idealized types of physical model bodies

Target	Transmission τ	Absorption α	Reflection ρ
Black body	0	1	0
White body	0	0	1
Opaque body	0	$\alpha + \rho = 1$	
Gray body	α, ρ and τ are uniform for all wavelengths		
Transparent body	1	0	0

In radiative transfer simulations or in simulations of measurement processes, the electromagnetic radiation emitted by natural media can be approximated by those of ideal bodies such as the black body discussed in Sect. 10.2.2. In contrast to a black body, a white body is completely and uniformly reflective in all directions. Another ideal body is the gray body, which presents wavelength-independent absorptivity. Table 10.3 summarizes some ideal bodies that can be taken into account.

The spectral characteristics of natural objects only correspond to those of above idealized bodies within limited wavelength ranges. As already discussed above for elastic scattering processes, different natural objects interact differently with incident solar radiation over the entire wavelength range due to differences in their physicochemical properties and surface structures. Reflection, absorption, transmission, refraction, double refraction, or more generally scattering as well as nonlinear optical properties are important influences

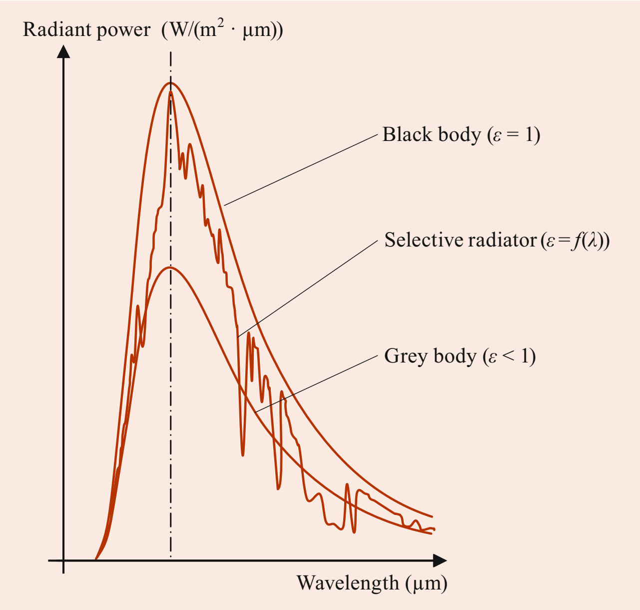


Fig. 10.3 Schematic illustration of the radiant power of a black body (10.4), a selective radiator (wavelength-dependent emissivity), and a gray body (wavelength independent emissivity) as a function of the wavelength radiant power

on the spectral signatures of natural objects, as will be explained in more detail in upcoming paragraphs. This means that the emissivity and absorptivity of natural objects can vary strongly and specifically with the wavelength. Therefore, natural objects are called selective backscatterers or selective emitters [6], and these objects can be identified and distinguished from each other based on their spectral characteristics (also called their *spectral fingerprints*). Figure 10.3 illustrates the wavelength-dependent emission of different radiators. Furthermore, these characteristic signatures enable material properties to be examined and qualitatively or quantitatively evaluated. By means of RS, simultaneous acquisitions of the same area in different wavelength ranges allow the differentiation of various materials and surface types.

Besides optical properties, thermal properties of natural media are also of interest in RS. Radiation emitted by objects is remotely sensed in the TIR range and deployed to detect and distinguish between different objects. An essential parameter to characterize the thermal behavior of an object is the thermal inertia b. This defines the ability of a material to resist a change in temperature when a periodic disturbance function is applied, and is a complex term that depends on the thermal conductivity k, specific heat c_H, and density ρ [1] such that

$$b = \sqrt{k\rho c_\mathrm{H}}\,. \tag{10.15}$$

Typical thermal properties of various soils and some of their components are shown in Table 10.2.

Atmospheric Processes

When electromagnetic radiation passes through the atmosphere, it interacts with atmospheric particles such as air molecules (e.g., N_2, O_2), aerosols (solid and liquid particles in the atmosphere), and cloud particles. The amount of interaction is determined by the types, numbers, and chemical components of the particles present in the path of the radiation, the length of this path through the atmosphere, and the wavelength of the radiation. Processes such as photoionization and the excitation of molecular vibrational and rotational states lead to absorption of some of the radiant energy. Other parts of the radiation are elastically scattered, depending on their size parameters, dielectric properties, particle shapes, and particle orientations with respect to the incident radiation in the case of nonspherical objects.

Since air molecules are very small (their diameters are typically below 1 nm), their size parameters are on the order of 10^{-2} to 10^{-3} in the near-UV and visible spectral range. Thus, the *Rayleigh approximation* (*Rayleigh scattering*, Sect. 10.2.4) can be applied to describe their elastic scattering behavior. For larger wavelengths, particularly in the microwave region, elastic scattering can be neglected.

Aerosol and cloud particles are, in general, much larger than air molecules (diameters typically range from 100 nm up to several mm), so the *Rayleigh approximation* is no longer applicable. Additionally, they are, in general, a mixture of different inhomogeneous and nonspherical particles, so more sophisticated models are needed to describe their elastic scattering characteristics. Examples of nonspherical particles include ice crystals in cirrus clouds and dry dust aerosols. Typical features of the scattering properties of nonspherical particles are enhanced side scattering compared to spherical particles, depolarization effects (spherical particles do not depolarize), and halo phenomena caused by hexagonal ice crystals.

However, these features are not observed or encountered in many RS scenarios. Thus, the assumption of spherical particles is often sufficient, within a certain error range, in the modeling of RS scenarios or in the evaluation and processing of RS data. As already discussed in Sect. 10.2.4, these particles can be treated within the Mie theory (Mie scattering).

In general, the particle ensembles are a mixture of objects with different sizes (radii) and chemical compositions (dielectric constants). To cope with this problem, various size distributions are considered for specific particle types. Integration over these distributions yields size-averaged scattering quantities. Inhomogeneous scatterers can be treated using, e.g., effective dielectric constants resulting from effective medium theories or from simple mixture formulas. Size-averaged scattering quantities of different particle types, characterized by special dielectric constants, can then be externally mixed to describe the whole ensemble or cloud.

Due to the abovementioned atmospheric scattering and absorption processes, signals from the Earth's surface are modi-

Table 10.4 Overview of possible influences of the most important atmospheric parameters (ozone, water vapor, molecule or *Rayleigh scattering*, aerosol scattering) on data of spectral reflectance and NDVI based on the example Landsat/ETM and NOAA/AVHRR. It shows maximum errors of corrected albedo as a function of variations of atmospheric state. Significant errors are highlighted in bold, italic; an "X" indicates large sensitivity, but without numerical assessment (after [28–30], cited in [31, 32]). D.U. - Dobson Unit, AOT - Aerosol Optical Thickness

Parameter (range of values, unit)	Ozone	Water vapor	Elevation	Aerosol content	Aerosol type	Surface type	Solar zenith
	(250–500 D.U.)	(0.5–4.0 g cm^{-2})	(0–2000 m)	(AOT_{550} = 0.05–0.8)	(absorption, scattering)	(anisotropic reflectance)	(0–70°)
Blue band (ETM1)			***+X***	***+X***	0.03 to **0.14**	0.04	< 0.01
Red band (ETM3, AVHRR1)	**−13.5%**	−4.4%	+0.07	**+0.12**	*X*		
Near-infrared band (ETM4, AVHRR2)	−0.5%	***−22%***	+0.04	+0.083	0.005 to 0.02	**0.12**	−0.05 to **+0.10**
Mid-infrared band (ETM5, ETM7)		***+X***			0.01 to 0.04	**0.30**	−0.05 to **+0.10**
Vegetation index for bare soil (NDVI = 0.05)	**+0.07**	**−0.12**	**−0.094**	**−0.085**	*X*		
Vegetation index for deciduous forest (NDVI = 0.85)	+0.017	−0.038	**−0.26**	**−0.34**		*X*	

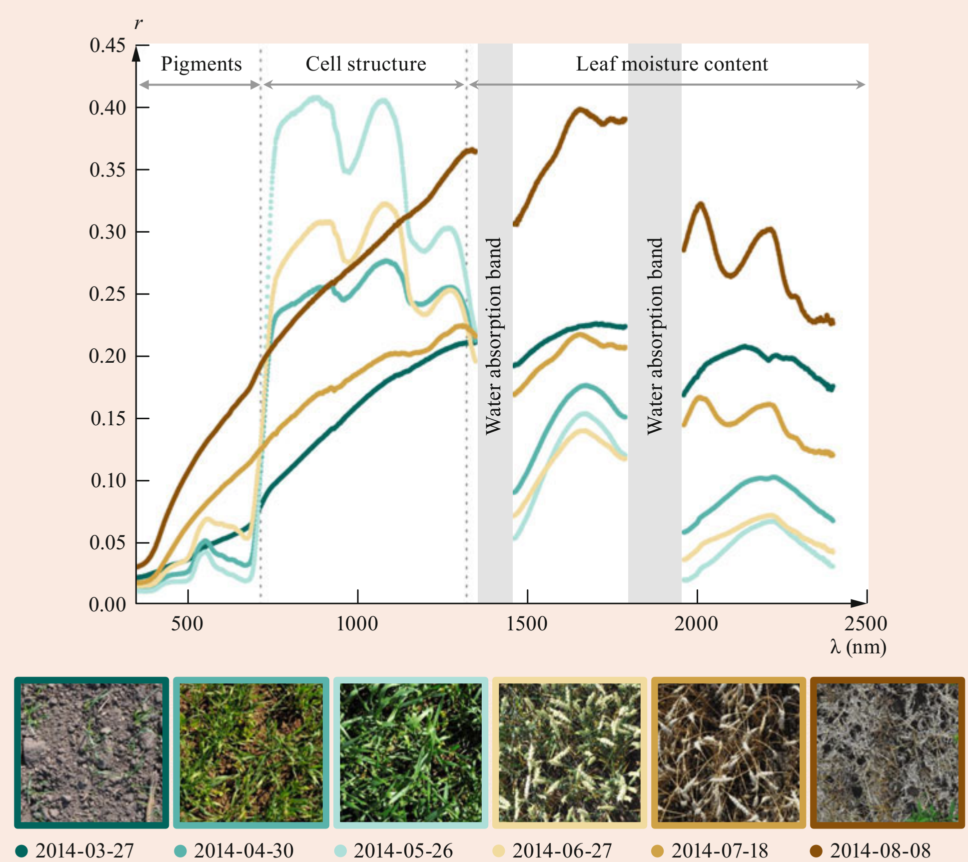


Fig. 10.4 Absolute reflectance factors (r) [33] representing different phenological stages of spring wheat growing on Haplic Chernozem at the Gebesee test site in 2014 [34]. The hyperspectral signal is sensitive to changes in the pigment content, the structure, and the moisture content of leaves in certain wavelength ranges [2], whilst changes in the non-vegetation-covered part of the soil and the soil moisture content influence the signal throughout the depicted wavelength range. Due to the influence of water absorption in the wavelength ranges from 1350 to 1460 nm and from 1790 to 1960 nm [35], the spectral reflectance factors are masked out

fied or even completely masked by the atmosphere as they pass through it before being measured at sensors. The larger the content of scattering/absorbing particles and air molecules, the stronger the effects. The degree of masking is also influenced by the geometry of the Sun–Earth's surface–sensor system. The most important absorption bands of the Earth's atmosphere with respect to the solar spectrum are those of water vapor, oxygen, carbon dioxide, ozone, and methane. In addition, aerosols and molecules cause a broadband attenuation of the solar radiation through elastic scattering [31].

Table 10.4 gives an overview of the possible influences of the most important atmospheric parameters (ozone content, water vapor content, molecular scattering (*Rayleigh scattering*), and aerosol scattering) as well as the directional dependence of the reflection at the ground and the solar–zenith angle on spectral reflectance data [28] and normalized differential vegetation index (NDVI) data from two sensors: the Landsat Thematic Mapper (TM) and the Advanced Very-High-Resolution Radiometer (AVHRR; [30, 36], cited in [31, 32]).

Interaction with Vegetation on Land Surfaces

The spectral remission that is characteristic of soils is generated in a complex way by the spectral characteristics of a combination of heterogeneous mineral, liquid, and organic components. The term *remission* refers to the diffuse reflection of light on non-mirror-reflecting objects. The most influential parameters are the source rock, soil minerals, organic matter content, water content, and surface roughness [1, 37]. The main difference between the remission characteristics of green plants and soils are the green peak around 550 nm, the steep increase at the transition to the NIR, and the persistence of this high remission level up to a wavelength of about 1300 nm (Fig. 10.4).

The spectral remission characteristics of green plant components are determined by various morphological, anatomical, and physiological factors [1]: (i) leaf surface characteristics (e.g., wax, roughness), which cause remission in the UV and blue; (ii) the composition, concentration, and distribution of leaf pigments (e.g., chlorophyll), which cause absorption in the UV and VIS ranges; (iii) the structure of the mesophyll tissue and the cell wall thickness, which cause remission in the NIR; and (iv) the water content, which causes absorption in the NIR. The remission characteristics of the plant canopy do not correspond to those of a single plant or individual plant components. Rather, they are determined in a complex way by its three-dimensional structure, which is influenced by the positions of the leaves, the remission of stems, flowers, and buds, the plant density, the shade in the stand, and the subsoil [1]. Additionally, plant populations show daily and seasonal variability, which affects the remission characteristics.

10.3 Technical Basics of Remote Sensing Data Acquisition

The term remote sensing is an abstract and comprehensive term that encompasses a multitude of different carrier platforms, sensors, and processing methods that can also be combined with each other. This results in a *multiconcept* [1].

RS systems can be differentiated according to diverse characteristics such as measurement type and spectral measurement region (Table 10.5; [38]).

10.3.1 Spatial, Spectral, Radiometric, and Temporal Resolution of Remote Sensing Systems

The suitability of a carrier–sensor configuration to solve a specific task depends essentially on the properties and dynamics of the investigated object or phenomenon and on the performance of the RS system. The performance of a RS system can be described by properties such as the spatial, spectral, radiometric, and temporal resolution (Table 10.5):

- The spatial resolution gives information about the size of the section of the Earth's surface that can be represented in one pixel. The term *ground resolution* refers to the ability of a sensor to detect two adjacent objects separately.
- The spectral resolution (bandwidth: mm) provides an indication of how well a RS sensor can distinguish different spectral ranges of the electromagnetic spectrum. The number of spectral channels of a sensor can be an indication of the spectral resolution.
- The radiometric resolution (unit: bit) determines how well a sensor can distinguish between intensity values.
- The temporal resolution (unit: hour or day) determines the time interval between individual images of the same area.

Besides this classification scheme, RS sensors are divided into passive and active systems according to their recording behavior, whereby wide ranges of the electromagnetic spectrum can be evaluated.

Passive systems record solar radiation reflected from the Earth's surface (e.g., multispectral scanners) and the radiation emitted by the Earth's surface (e.g., a thermal imaging camera). In contrast, active systems emit microwave or laser beams and receive their reflected components (e.g., radar systems and laser altimeters).

Table 10.5 Taxonomy of RS systems (adapted from [38])

Taxonomy of remote sensing systems							
Recording platform	Satellite/shuttle	Aircraft/balloon	Small planes/UAVs	Stationary			
Recording mode	Imaging mode	Nonimaging mode	Sounding mode				
Recording behavior	Passive (VIS, NIR, TIR, thermal MW)	Active (lidar, radar)					
Recording type	Analog (camera, video)	Digital (whiskbroom, line array, 2D CCD)					
Spectral coverage	Visible/ultraviolet	Reflected infrared	Thermal infrared	Microwave			
Spectral resolution	Panchromatic 1 band	Multispectral 2 to < 20 bands	Superspectral 20 bands	Hyperspectral > 20–250 bands	Ultraspectral > 250 bands		
Radiometric resolution	Low (< 6 bit)	Medium (6–8 bit)	High (8–12 bit)	Very high (> 12 bit)			
Spatial resolution	Ultralow > 500 m	Very low > 250 m	Low 50–250 m	Medium 10–50 m	High 5–10 m	Very high 1–5 m	Ultrahigh < 1 m

RS systems can also be categorized into imaging, nonimaging, and sounding systems based on the types of measurements performed. Imaging systems measure the intensity of the detected radiation as a function of the position on the Earth's surface and convert this into a 2-D intensity image. Nonimaging systems either measure the radiation intensity but not the position on the Earth's surface or vice versa. Sounding systems (sounders) measure the intensity of the radiation to provide information about a particular property as a function of height in the atmosphere.

10.3.2 Passive Sensors

Passive remote-sensing systems detect energy that is reflected or emitted by the observed object (such as reflected sunlight, emitted thermal or microwave radiation, and night lights); the energy never originates from the RS system itself. The most important radiation sources for passive sensors are the Sun and the Earth.

Photographic Remote-Sensing Systems

In the past, photogrammetric cameras were the sensor systems most frequently used for RS. An aerial photograph is an image of the Earth's surface taken with a camera that is mounted on a flying platform. Initially, those platforms were balloons; aircraft and (rarely) helicopters have since been used as well. Drones have also become popular platforms for small-scale projects. Photogrammetric data capture makes use of so-called aerial cameras that enable the recording of a series of consecutive aerial photographs during the photoflight.

In the era of analog cameras, black and white (panchromatic, IR), color, and false color films were available. Since the turn of the century, digital aerial cameras have been predominant, and multi-camera-head approaches are typically employed, where the additional heads are normally used to separately record multiple spectral bands. In addition, images from several camera heads are combined (*stitched*) to obtain large-format (virtual) images. Some of the latest digital cameras only use RGB sensors, meaning that there is no need for subsequent pan-sharpening.

Aerial photographs are commonly classified according to the orientation of the optical axis. Near-vertical images (taken $< 3°$ from the vertical) present the terrain in a similar manner to a map. Oblique images (taken about 45° from the vertical) present a birds-eye view of the terrain, and are widely used for building 3-D city models because they also show the facades of buildings, not only the roofs.

Depending on the lenses and the sensor size, aerial cameras can be classified into wide-angle and normal-angle cameras with a field of view (FOV) of about 90° and 60°, respectively. The use of a wide-angle camera leads to a more stable geometry of the photogrammetric block because of better (near to 90°) ray intersections, while a normal-angle camera delivers better ground views in narrow urban environments. However, modern digital cameras offer a range of different sensor sizes and lenses with different focal lengths. Thus, several FOVs are used today.

For economic reasons, aerial cameras are generally large-format cameras. Historically, they had a standard film format of 23×23 cm, whereas modern aerial cameras provide large digital images up to 375 Mpix (DMC III, pan, $25\,728 \times 14\,592$ pixels) and 450 Mpix (Ultracam Eagle Mark 3, pan, $26\,460 \times 17\,004$ pixels, stitched) in size. For area-wide mapping and surveying purposes, photographs are taken in strips, with an along-strip overlap of 60%, 80%, or even more for consecutive images and an across-strip overlap of 15–35% historically and up to 60% today.

Since the turn of the century, digital aerial cameras have replaced analog cameras, so the latter are no longer in use. However, the data recorded by aerial survey cameras in the past are increasingly being included in time series analyses. Therefore, images on film are still relevant.

Television (TV) Cameras for Remote Sensing

TV technology such as the return-beam vidicon (RBV), focus projection and scanning (FPS) vidicon, direct beam readout vidicon, image dissection camera, dielectric tape camera, secondary electron conduction (SEC), and the electron-bombarded silicon (EBS/SIT) tube were used onboard various satellites such as Landsat 1, 2, and 3 [39]. The RBV camera system produced high-resolution television-like images of the Earth's surface. Its design was similar to that of conventional photographic camera technology, but the film in the focal plane was substituted for a photosensitive plate [40].

The solar radiation reflected from the Earth's surface was focused by a lens onto a photosensitive plate in order to expose it. Depending on the incident light intensity, the sensor elements changed their charge states. These states were then read out using an electron beam and converted into a digital signal (A/D conversion), which was transmitted to a ground station [40]. The resolution of this method is essentially determined by the diameter of the recording electron beam. TV and photographic procedures work in both the VIS and the NIR ranges. Recording multispectral data requires several RBVs and upstream filters.

Scanners

Scanners are optical sensors that can be used to detect landscape objects on the Earth's surface in the spectral range from the UV to the TIR (0.3–14.0 μm). Scanners of various designs are deployed in RS, such as electromechanical scanners (operated on TIROS, NIMBUS, Landsat 1 and 2) and optoelectronic scanners (e.g., the HRV mounted on SPOT-1). The flight line on which a satellite moves around the earth is its orbit. If a scanner captures an image area across to the orbit, a wide scanning strip (also known as a swath) is scanned on the earth's surface. If the earth now rotates away from under the satellite orbit, a complete coverage of the earth surface can be realized. Another scanner concept is the conical scanner (e.g., MSU-SK 1 mounted on RESURS 1–3; [39]).

The radiation reflected and emitted by the target objects is detected by an optical system and focused on detectors. The detectors convert the incident electromagnetic radiation into electrical signals. Before the data is stored and transmitted to a receiver, it is amplified. A multispectral scanner records the same terrain in several partial spectral ranges simultaneously. The images are congruent since spectral decomposition of the radiation only takes place after it has passed through the optical system.

The spatial resolution of a scanner is essentially determined by the sensor flight altitude h, the spatial resolution of the optical system, and the number of detectors. The spatial resolution is given as either the angular resolution AR (in mrad) or the resolution length (in m). The resolution of a scanner image is primarily determined by the size of the scanned ground elements, which depends on the sampling frequency, rotation frequency f_{h}, the path speed, and the opening angle o of the optics. The distance from the sensor to the ground elements and overlap of the scanned ground elements also affect the ground resolution. The distance to the ground elements depends on the speed of the sensor carrier, the flight altitude, the sampling frequency, and the opening angle of the scanner [11].

The IFOV (instantaneous field of view) can be computed using

$$\mathrm{IFOV} = \mathrm{AR} \cdot h \,. \tag{10.16}$$

In order to distinguish terrain objects from each other, these objects must be larger than the IFOV or, if the objects to be imaged are smaller than the IFOV, they must provide particularly high radiation contrast compared to their surroundings. The spectral resolution of scanners depends on the number and the type of detectors used. The thermal noise in the detectors influences the radiometric resolution.

Optomechanical Scanners (Whiskbroom Scanners)

Optomechanical scanners are also referred to as across-track scanners. The components of these systems are a mechanical scanning system, an optical system, and a detector unit. To record a section of the Earth's surface, the radiation energy scattered from the Earth's surface to the sensor is focused onto the detector via a mirror system and an optical system. By rotating or tilting the mirror, the ground elements can be imaged together perpendicularly to the flight direction to form a strip. Afterwards, the mirror returns to its original position and generates the next strip. The mirror movement is synchronized with the airspeed in such a way that the recorded data is aligned in the direction of flight [39].

Optoelectronic Scanners (Pushbroom Scanners)

Optoelectronic scanners are also called along-track or pushbroom scanners. In contrast to electromechanical scanners, optoelectronic scanners have no susceptible mechanics (e.g., a mirror or shutter). The recording principle for this type of scanner is based on the simultaneous recording of a strip of the Earth's surface by a detector line. The image is taken from the central perspective and has no panoramic distortions. The radiation impinging on a detector is integrated over a time period and converted into a readable signal. The detector line located in the focal plane of the optics is read out sequentially. Individual detectors of a scan line are thus recorded simultaneously [39].

The distance traveled over the Earth's surface due to the forward movement of the satellite during the integration time defines the ground resolution in the flight direction. The ground resolution orthogonal to the direction of motion is defined by the number of detectors (e.g., CCDs) and the optics used.

To record multispectral data, several detector lines must be arranged next to each other. If these lines record different areas of the electromagnetic spectrum, spectral channels are obtained. Due to the very good signal-to-noise ratio (SNR) of CCDs, these detectors provide high radiometric resolution and are preferentially used in Earth observation.

Conical Scanners

Scanner systems operating in the nadir direction often scan large circular arcs that intersect with the subsatellite orbit vertically. The geometric conditions that prevail here are relatively simple. The area of scanned ground elements increases with their distance to the nadir point. This effect, called panorama distortion, causes an increase in the radiant power received per ground element. Radiometric corrections are necessary to overcome this distortion. The geometric conditions of systems operating in the nadir direction have been extensively described in the literature [1, 3].

In the case of conical scanners, the geometric conditions are more complicated than nadir scanners. The conical scanner is located at the apex of the cone axis. The cone axis is pointed towards the nadir, and only a forward-looking circular segment is recorded. Data acquisition is carried out with a constant viewing angle, so that the scanning describes the shell of the cone. Therefore, all ground elements are of the same size, which has positive consequences for the radiometric ratios of the image data. Furthermore, the ground elements lie on a small circle on an arc of the terrain surface recorded by the scanner. As a rule, these ground elements are configured symmetrically to the flight direction of the satellite, and the number of ground elements scanned is always the same [41].

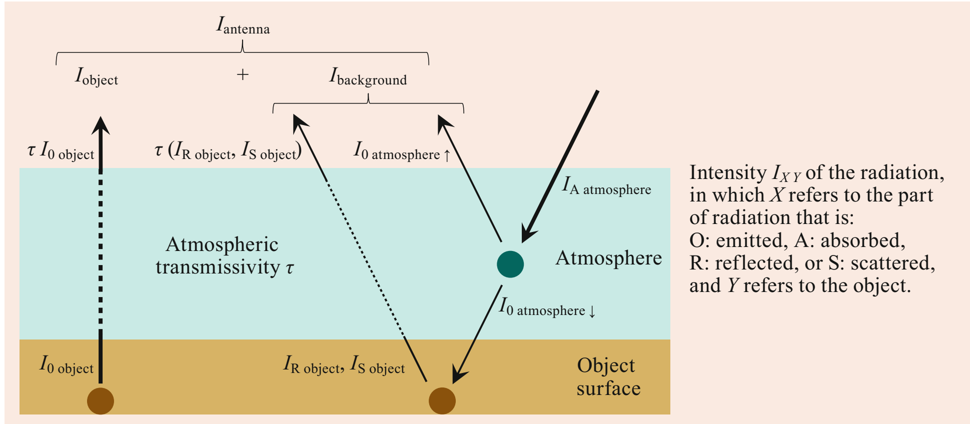


Fig. 10.5 Composition of the radiation received by the antenna of a radiometer [14, 42]

Radiometers and Scanning Radiometers

Radiometers are passive sensors enabling the remote detection of radiation that is emitted by objects (with $T_{\text{object}} > 0\,\text{K}$) and occurs as a consequence of the movement of atoms and molecules in the objects [14, 42]. The radiant exitance $M_{\text{object}}(\lambda)$ of such an object depends on the temperature T_{object} and the emissivity ε' of the object surface and can be quantified by the Stefan–Boltzmann law (10.9) [43]. The radiation emitted from an object is influenced by the transmissivity τ of the media (e.g., the atmosphere) it passes towards the radiometer. When interpreting the measured signal, it should also be kept in mind that the radiation can comprise components that originate from upward-directed atmospheric emissions ($I_{0\text{ atmosphere}\uparrow}$), are reflected ($I_{\text{R object}}$), or are, in general, scattered ($I_{\text{S object}}$) at the surface of the object (Fig. 10.5 [14, 42]).

The basic operational principles for radiometers are the same in the infrared and microwave ranges [43]. Nevertheless, we should distinguish between infrared and microwave radiometers that operate in the wavelength ranges from 8 to 14 µm [14] and from 0.3 to 6 cm [5] for the following reasons:

(i) For various types of surfaces, the characteristics of the emissivity ε vary according to the investigated wavelength range [43]. For example, in the infrared range, the emissivity ε for the sea is close to 1, and it is more or less unaffected by changes in the properties of the medium for diverse land-cover types [3]. In the microwave range, the emissivity for the sea is still almost constant but deviates from 1, while the emissivity changes to an increasing degree depending, for instance, on the ability of the land cover type to absorb water (thereby changing its electrical conductivity), among other factors [19, 43].

(ii) The intensity I emitted at the Earth's surface is approximately 10^5 to 10^6 times higher in the infrared range compared to that in the microwave range [5]. Therefore, microwave radiometers are equipped with highly sensitive receivers in the form of parabolic or horn antennas that generate almost no noise power [5, 42]. Furthermore, they cover a large bandwidth, and the IFOV is 3–4 times larger than that of an infrared radiometer [43] to also allow for small integration times.

(iii) Microwave radiation has longer wavelengths than infrared radiation, which enables the former but not (in general) the latter to penetrate clouds [43]. For this reason, the applicability of microwave radiometers—for instance their utilization to monitor snow cover, soil moisture, or oil pollution [3]—is almost independent of the prevailing weather conditions [43]. In contrast, the monitoring of sea or land surfaces with infrared radiometers is limited to cloud-free conditions, although they are deployed to investigate in-cloud temperatures [14].

(iv) In contradistinction to infrared radiometers, microwave radiometers can capture horizontally and vertically polarized radiation separately [43].

10.3.3 Active Sensors

In contrast to passive RS systems, active RS systems illuminate the area under investigation on the Earth's surface and record the reflection signal. Therefore, the signals recorded by these systems are independent of solar radiation, and data can be recorded during the day and at night. There are two functional principles of active sensors. These are radar (radio detecting and ranging) sensors, which operate at frequencies in the microwave range (see Table 10.1), and lidar (light detecting and ranging) sensors, which operate at frequencies in the near-UV, visual, near-infrared, and mid-infrared regions of the electromagnetic spectrum.

Radar Sensors

The antenna of an imaging radar system is oriented with its longitudinal axis in the direction of flight and its transmission and reception direction perpendicular to the flight path diagonally downwards (side-looking airborne radar: SLAR). This configuration is designed to prevent all the backscatter signals from the transmitted radar impulses from being received simultaneously. The terrain is scanned in strips by the radar signal transmitted by the antenna, looking across the flight direction [44]. SLARs can be categorized according to their main functional principle for imaging the Earth's surface into real aperture radar (RAR) and synthetic aperture radar (SAR) systems [5]. Aside from an antenna, radar systems comprise a transmitter, a receiver, a switch for the antenna between transmission and reception modes, and a storage unit for recording the signals.

Depending on the respective wavelength, microwaves can be used for the detection and tracking of rain clouds (e.g., K_a to W band) or to penetrate atmospheric cloud and rain phenomena almost interference-free (e.g., X to P band). For these systems, the day/night change does not affect the signal, and atmospheric effects play only a minor role in the long-wavelength range. With these systems, the Earth's surface can be recorded in the same way during both the ascending and descending modes.

Due to the use of different frequency ranges from those employed in optical RS, the type of imaging performed, and the use of a coherent transceiver system, radar images differ in many aspects from optical images. Therefore, an understanding of the characteristics of radar images is required for the interpretation of, e.g., the SAR image content.

The geometry of radar images is dominated by the recording method, the oblique view from the sensor of the terrain surface, and the terrain shape. The oblique view leads to geometric distortions with respect to the ground geometry. In addition, the oblique view of uneven terrain causes effects such as foreshortening, layovers, and radar shadows. These effects also have a radiometric effect on the image. The geometric and radiometric distortions cannot be eliminated completely by conversion into the ground plane or by georeferencing.

The foreshortening effect results from the fact that the terrain inclined towards the radar sensor in the oblique view is reached almost simultaneously by the radar impulse front, so the different reflection signals reach the radar at nearly the same time. Therefore, these terrain forms are geometrically shortened. The affected area is radiometrically brightened.

Very steep angles of incidence lead to a layover effect, since the upper edge of a terrain or house wall, for example, is reached earlier by the radar impulses, causing them to be backscattered earlier than those at the foot of the terrain or house wall. The result is a superposition of different signals whereby layover areas appear radiometrically very bright.

Radar shadows are caused by terrain areas with inclinations greater than the depression angle of the radar beam. Therefore, these terrain areas are not illuminated by the radar, so they do not provide backscattering. In the image, these areas appear radiometrically dark.

The penetration depth of microwaves into the terrain surface is generally significantly greater than the penetration depth of optical radiation. Depending on the wavelength, the direction of radiation incidence, and the properties of the terrain surface layer, it ranges from a few millimeters to a few decimeters. In particular, microwaves can penetrate the vegetation layer and thus provide information about the subsurface.

Due to the forward movement of the sensor, area-wide detection of the terrain surface is realized. The intensity, transit time, and polarization of the transmitted signal are measured. Table 10.1 gives an overview of the frequency ranges of radar systems.

Real Aperture Radar (RAR)

When considering the spatial resolution of radar systems, we can distinguish between the along-track (or azimuth) direction, which depends on the antenna aperture, and the cross-track (or slant range) direction, which depends on the pulse length [5, 44–46].

The spatial resolution of a real aperture radar (RAR) system is limited in the azimuth direction by the directional characteristic of the antenna or the half-width of the main lobe of the antenna, which is significantly influenced by the antenna size [46]. The azimuthal beam width of the antenna is the ratio of the wavelength of the emitted radiation λ to the antenna length l [5, 44, 45],

$$\beta \cong \frac{\lambda}{l} \qquad (\mathrm{rad})\,. \tag{10.17}$$

The azimuth resolution at the terrain surface R_a can be calculated from the azimuthal beam width β and the distance R between the object and antenna (i.e., from the wavelength of the emitted radiation λ, the antenna length l, the height of the platform h, and the look angle ϕ) via

$$R_a \cong \beta R = \frac{\lambda}{l} R = \frac{\lambda}{l} \frac{h}{\cos\phi}\,. \tag{10.18}$$

The resolution in the flight direction essentially depends on the duration of the radar pulse τ and the velocity of the electromagnetic radiation c.

$$R_r \cong c\frac{\tau}{2} \tag{10.19}$$

The denominator of 2 in the equation results from the fact that the impulse must travel twice the antenna–object distance. Short pulses provide higher resolution. Assuming a classical RAR system, a wavelength of 5.65 cm, an antenna length of 12 m, and an assumed antenna–target distance of approximately 800 km, the resolution in the azimuth direction would be about 3.8 km. If an azimuth resolution of 5 m was required, the antenna length would have to be 9 km.

10

Practical limits on the real aperture size are clear from (10.18). Payload restrictions present additional problems. Therefore, RAR sensors are not suitable for satellite applications due to their poor spatial resolution. To overcome the limitations of a RAR, the SAR concept was developed.

Synthetic Aperture Radar (SAR)

In synthetic aperture radar (SAR), the radar echoes received from a ground object by the antenna at different positions along the flight path are collected and offset against each other. The distance passed during the time that the ground object lies within the directional beam corresponds to the length of the synthetic array. During the movement of the antenna, each object is illuminated and recorded with a variable viewing angle. If the path of the small real antenna is known with sufficient accuracy, the aperture of the larger antenna can be reconstructed from the intensity and phase of the received radar echoes [46]. This leads to a better ground resolution for SAR than for RAR. The length of the synthetic antenna L is the flight distance during which a certain terrain object remains in the scanning range of the short real antenna. This distance is also identical to the spatial resolution R_r of the real antenna [45].

In their analyses of the azimuth resolutions for SAR systems, [44, 46, 47] come to the conclusion that the best possible theoretical resolution in the azimuth direction R_a of SAR is half the real aperture length. Thus, the spatial resolution of SAR in the flight direction is given by

$$R_a = l/2 . \qquad (10.20)$$

A small antenna length l therefore leads to a better azimuth resolution R_a, as the half-width of the antenna lobe and thus also the length of the synthetic aperture is increased. However, the improvement in the spatial resolution achieved by reducing the aperture is limited because the antenna loses sensitivity as its dimensions decrease. Table 10.6 explains some of the technical terms used in connection with radar remote sensing [46].

A SAR system generates images by coherently processing the backscattered signals. These systems are therefore susceptible to the so-called speckle effect. This is granular noise that can be traced back to the interference of waves reflected by many elementary scatterers. This interference causes phase differences between the waves. The speckle effect is therefore a special feature of radar images [48, 49].

Since this form of noise reduces the quality of radar images, thus making it difficult to evaluate the data (e.g., via segmentation or classification; [49]), various methods have been developed to minimize the speckle effect. There are different methods of attenuating a speckle pattern, e.g., filtering or multilook processing, but this always means a certain loss of information.

Radar-Altimeters

Radar altimeters are nonimaging radar systems. These RS systems provide the distance (altitude) between the respective carrier (e.g., a satellite or aircraft) and the Earth's surface by transmitting an impulse and measuring the time between transmission and receipt of a backscattered ground track signal of the topographical profile. The height h is linearly connected to the time t [50] via

$$h = (ct)/2 . \qquad (10.21)$$

Altimeters were included in the payloads on the satellites SeaSat, ERS-1, ERS-2 (ERS refers to European Remote Sensing), and the TOPEX/Poseidon mission [39].

Radar Scatterometers

Scatterometers are nonimaging radar systems that emit electromagnetic energy towards a target object in order to quantitatively measure the scattering coefficient. Scatterometers were included onboard various satellites such as SeaSat-A, ERS-1, ERS-2, MetOp-A, and MetOp-B as payloads. A C-band scatterometer was used on the ERS satellites to measure the wind speed and wind direction over oceans. Additionally, the system was used to extract the soil moisture (wetness index) of the uppermost soil layer (0.5–2 cm deep) by performing measurements in the C band. An overview of scatterometry can be found in [51].

Table 10.6 Basic terms and definitions associated with synthetic aperture radar RS (adapted from [46])

Basic term	Definition
Trajectory	Flight path of the carrier platform
Azimuth/along track	Direction of the carrier platform or antenna in the direction of movement
Range/cross-track	Direction transverse to the direction of movement of the beam
Footprint	Surface area currently illuminated by the radar
Swath	Total area of the terrain surface that is covered by the antenna as a result of the movement of the platform
Radar lobe	Main maximum of the far-field radiation pattern of the radar antenna
Slant range	Distance from the antenna to the target in view
Ground range	Projection of the antenna–target distance onto the Earth's surface
Nadir	Base point of the current position of the carrier platform on the Earth's surface
Baseline	Distance between locally separated transmitter and receiver antennas (local separation can mean that these antennas are mounted on different platforms or on the same platform a certain distance from each other)

Lidar Sensors

Analogously to radar systems, lidar (light detection and ranging) systems also use the so-called echo principle, but in the ultraviolet, visible, and near-infrared wavelength ranges (radar systems operate in the microwave range). A lidar uses laser light to measure the distance between an object and the sensor and to gain information on the object. The laser generates highly focused and short scanning beams (e.g., a few mrad) of, in general, high-intensity monochromatic light with a long coherence length, and sends those beams towards the object. The energy is emitted continuously or in short pulses with a high repetition frequency. Scattered signals reaching the sensor are captured by telescope optics and converted into electrical signals by detectors.

Due to the wavelength ranges used, the transmitted pulses can be affected by atmospheric phenomena such as refraction or dispersion, so it may be necessary to correct the signals.

The main applications of lidar are measurements of distance (height), motion, and chemical composition. By measuring the intensity, polarization, and spectral properties of the reflection signal as a function of time, it is possible to obtain information about the properties of the atmosphere, vegetation, and the Earth's surface. This is in large part achieved by overcoming the technical problems associated with spaceborne lidar systems. It is now possible to determine profiles of aerosols and clouds (e.g., Cloud-Aerosol lidar and Infrared Pathfinder Satellite Observation (CALIPSO), launched April 2006; [52, 53]; Earth Cloud Aerosol and Radiation Explorer (EarthCARE), launch planned for 2023; [54]), to measure vertical wind profiles (AEOLUS, launched in 2018; e.g., [55]), and to obtain spatial and temporal gradients of atmospheric methane columns (Methane Remote Sensing Lidar Mission (MERLIN), launch planned for 2027; e.g., [56]) using lidar systems in space.

10.3.4 Earth Observation

Satellite-based Earth observation is a subarea of RS (Sect. 10.2.3) that involves recording data and information on the physicochemical and biological properties of this planet's systems using RS technologies. Earth observation is used to monitor and evaluate the conditions of and changes in, for example, natural and urban environments. In contrast to other RS strategies, it is characterized by several special features that must be met, such as various orbits, carrier systems, and higher costs for data acquisition.

The term *Earth observation* includes a space segment (e.g., sensors, satellites, platforms, and energy and communication modules) and a ground segment (operating equipment for the space segment, data reception and storage, preprocessing, sales and distribution, and data processing for the application-specific information carrier or image product).

Table 10.7 Taxonomy of satellites according to mass [57]

Satellite size category	Satellite mass (kg)
Large	>1000
Medium-sized	500–1000
Mini	100–500
Micro	10–100
Nano	1–10
Pico	0.1–1
Femto	<0.1

The space segment of a spaceborne RS mission can be divided into the launcher, space infrastructure, and satellite. One part of the satellite is the payload, which can be defined as the load onboard that is required to perform the specific task of the mission (e.g., sensors, instruments). Different payloads can be distinguished according to their importance for the objective of the mission. A primary payload serves to carry out the main task. A secondary payload acquires additional measurement data or performs other subordinate tasks.

Space Segment

The space segment of a RS satellite comprises a satellite or a satellite constellation (this may, in some cases, include spare satellites that can be exchanged for defective or decommissioned satellites), the sensor and/or instrument or a package, and the uplink and downlink satellite links. Following the ground segment, the remote sensing data will be distributed to users (e.g., research, industry, services) who will further refine them for their respective applications. Satellites can be divided into different size categories (see Table 10.7 [57]).

Satellite-based RS utilizes different orbits to perform different tasks depending on the Earth–satellite constellation. Orbits can be either circular or elliptic. A detailed description of the orbits utilized for Earth observation is given in Table 10.8.

In the context of Earth observation, an orbit is defined as track of a satellite around the Earth's surface. A satellite orbit can be described by various properties, e.g., altitude, inclination, and shape. The altitude of a satellite is defined as distance of the satellite above Earth's surface. The inclination of a satellite orbit is the angle of the orbital plane with respect to the equatorial plane.

In addition to these, parameters such as the ground track, zenith, nadir, swath, sidelap, and overlap can be used to characterize a RS system. The ground track is the vertical projection of the satellite orbit on the Earth's surface. While the satellite moves, the nadir also moves. The sequence of the nadir points defines the ground track. Every point in the direction above the satellite that is directly opposite the nadir is called the zenith. The nadir is the point on the Earth's surface that lies vertically below the satellite/sensor at the shortest distance from it. The swath of a satellite is the width of the

Table 10.8 Characteristics of the orbits utilized for operational Earth observation missions with various environmental research aims [39, 58–65]

Type of orbit	Abbreviation	Altitude 10^3 (km)	Orbit time	Characteristics	Advantages	Disadvantages
Geostationary Earth orbit	GEO	≈ 36	24 h	Located: above Earth's Equator (inclination i with respect to the equatorial plane is 0°) Illumination conditions: Varying Application: Meteorological satellites Special feature: satellite on GEO is equal to the rotational speed of Earth (eccentricity $\varepsilon = 0$)	Image recording	Data
Sun-synchronous orbit	SSO	0.5–1.5		Located: satellite passes over the polar regions; orbital inclination i = 97°–102°, Illumination conditions: data is always recorded at the same solar altitude and illumination conditions, Application: Earth observation, meteoroogical monitoring, global mapping Special feature: Sampling frequency depends on the distance between two consecutive orbits and swath width of the sensor; (equator: lowest, poles: highest)	Polar orbit allows: global monitoring of Earth with one satellite, multi-temporale observation; data of different times are comparable, because the illumination conditions high spatial resolution due to low orbital altitude, Uniform sampling resolution along the satellite's ground track due to the circular orbit, which results in constant satellite speed	No consideration of seasonal illumination variations possible Data cannot be collected with high time resolution by only one satellite Data on neighbouring swads are recorded on time
Dawn-to-dusk orbit (a special Sun-synchronous orbit)	Dawn-Dusk			Satellite crosses Equator around sunrise or sunset. Application: SAR satellites	Permanent exploitation of solar energy, since the satellite's solar panels are always illuminated by the Sun	
Low Earth orbit	LEO	0.2–1.5	90–120 min	Located: low Inclination: e.g., space station; crossing nrorth pole: e.g., military missions Illumination conditions: Varying Application: Manned space flight, e. g., International Space Station (ISS), Earth observation, meteorological satellites Special feature: Easiest reachable, most energy-efficient orbit, Spacecraft moves with a speed of 7 km/s	Multiple radio contacts with a ground station per day. High data rates	Short radio contact (10–15 min) with a ground station per orbit
Elliptical Earth orbit	EEO	≈ 0.1/ 0.3–40	12 h	Located: Orbital inclination i = 63.4° Illumination conditions: Varying Application: communication satellites Special feature: High eccentricity; temporal quasi-stationary observations of hemisphere	Satellite spends 2/3 of orbital period over one hemisphere	

area on the surface of the Earth that is imaged by the sensor during a single pass.

In space travel, different types of orbits can be distinguished, including a low Earth orbit (LEO), a geostationary Earth observation orbit (GEO), an inclined geosynchronous orbit (IGSO), a geosynchronous transfer orbit (GTO), a Sun-synchronous orbit (SSO), a medium Earth orbit (MEO), a highly elliptical Earth orbit (HEO), a halo orbit, and an elliptical Earth orbit (EEO). A detailed description of these orbits is given by [39]. Table 10.8 lists the types of orbits along which the majority of RS satellites are traveling (Fig. 10.6). Figure 10.6 shows the different types of orbits with their specific features used for RS applications.

Ground Segment

"The ground segment is the natural link between the Earth observation sensor in space, its operation service and the user on the ground" [65]. Major elements of an Earth observation ground segment include the flight operations segment/mission operations system (FOS/MOS), the payload ground segment (PGS), and the instrument operations and calibration segment (IOCS).

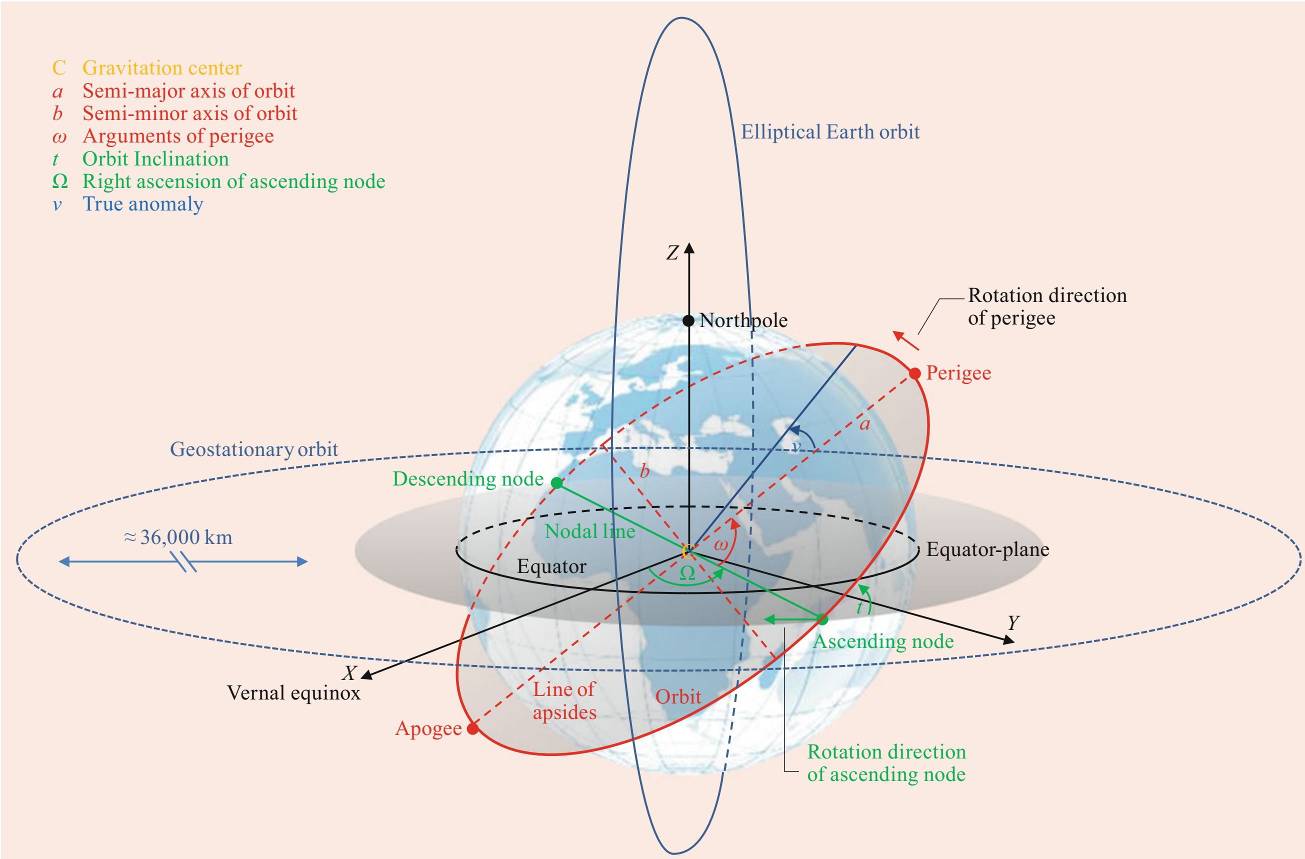


Fig. 10.6 Diagram showing different types of orbits traveled by RS satellites and various elements of an orbit

The mission operations segment (MOC) is a facility that is responsible for mission planning and for commanding the spacecraft and its instruments. The MOC monitors the health and safety of the spacecraft, verifies and validates the satellite and the payload components according to defined system parameters, and plans orbital maneuvers (such as those that place and keep the satellite in the correct orbital position and place the satellite into its final configuration for customer operations).

The instrument operation and calibration segment (IOCS) comprises functions such as instrument operation and control, internal calibration, external calibration, and the coordination of these tasks.

The payload ground segment (PGS) serves as interface for the data signals to and from the space segment. The PGS is responsible for, e.g., handling the payload data (processing, archiving, and cataloging) and product distribution. Both of these are independent elements of the ground segment that strongly interact with each other [65]. The payload ground segment receives the raw data and the (preliminary) auxiliary data (e.g., calibration data, housekeeping data, orbit prediction files). In most cases, payload ground stations are integrated into networks with international partners so that comprehensive coverage can be achieved.

Years ago, satellites did not have sufficient onboard memory to store payload data, so the data had to be downloaded in real time as the satellite entered the visibility circle of a ground station. Today, satellites are normally able to store all measured data onboard and to transmit the data to a ground station if the satellite is in its reception range.

However, in order to achieve greater flexibility in data transmission and to ensure that the ground segment has real-time capabilities (given the increasing number of sources delivering data packages of different sizes onboard satellites), the space packet protocol developed by the Consultative Committee for Space Data Systems (CCSDS) is currently used for telemetry transmission.

Different frequency bands, e.g., the S, X, and K_a bands, are used for the transmission of RS data. The lowest transmission rate is achieved with the S band, and the highest with the K_a band. Commands and housekeeping data are transmitted and received via the S band, whereas raw image data is downloaded via the X or K_a band.

Since data transfer between the ground station and satellite is affected by a wide range of environmental and technical influences, certain requirements are placed on ground stations, such as a visibility of 5° elevation above the horizon, an availability of 24 h/365 d, a reception security of $> 98\%$ of the requested data, or a bit error rate of less than 10^{-6}.

10.4 Data

RS sensors capture radiation reflected or emitted by elements and objects that constitute the land surface, ocean, and atmosphere. The characteristics of the resulting images are influenced by physical parameters such as intensity, runtime, wavelength, spectral information, polarization, phase, spatial information at the macroscopic level, temporal information, and angular information [4]. Accordingly, the sensors do not directly record the physical and chemical properties that are required by the user or model [66].

To harness the raw data, a basic understanding of the different data formats generated during the recording and processing of RS data is required. RS data are created and processed as two- or three-dimensional raster data. They are accompanied by metadata or housekeeping data that are created and delivered in ASCII or XML format (Sect. 10.4.1). The information required by the user or model can be extracted either from the RS data only (such as precipitation, air temperature, and vegetation indices) or from a combination of RS data with additional and/or in situ data (such as yields, hydrological data, and crisis information). These additional and in situ data, which are available in raster or vector format, are characterized in Sects. 10.4.1, 10.4.3, and 10.4.4, respectively.

The raw RS data are the actual payload measurement data from the main payload sensor (Sects. 10.3.2 and 10.3.3) onboard the carrier [67]. Because different sensors have different characteristics, RS data vary in spatial, spectral, and radiometric resolution and with regards to the area covered at the Earth's surface (the swath).

Different data product levels have been defined in the literature (ISO/TS 19159-1, [68]). Since the aspects of the various level definitions are not always comparable, the user should check the processing steps of the product of interest in detail. To ensure comparable processing of data originating from different carrier platforms and sensors, the ISO/TS 19159-1 *Calibration and validation of remote sensing imagery sensors and data* defines the following product levels:

- *Raw data*: Unprocessed original data.
- *Level 0 (L0)*: Reconstructed unprocessed data at full space-time resolution with all available supplemental information to be used in subsequent processing (e.g., the ephemeris and health and safety information) appended.
- *Level 1 (L1)*: Reconstructed unprocessed data at full resolution, time-referenced and annotated with ancillary information including radiometric and geometric calibration coefficients and georeferencing parameters (such as the ephemeris) that have been computed and appended but not applied to the L0 data. Radiometrically corrected and calibrated data in physical units at full instrument resolution as acquired.
- *Level 2 (L2)*: Derived geophysical parameters (e.g., the sea surface temperature and surface reflectance) at the same resolution and location as L1 source data.
- *Level 3 (L3)*: Data or retrieved geophysical parameters (e.g., the leaf area index) that have been spatially and/or temporally resampled (derived from L1 or L2 products), usually with some degree of completeness and consistency. Such resampling may include averaging and compositing.
- *Level 4 (L4)*: Model output or results from an analysis of lower-level data (these are parameters that are not measured directly but are derived from instrumental measurements).

10.4.1 Metadata

Parallel to the acquisition of RS data, other data describing the actual RS data (for example, the geographical location of the recording and the RS data quality) and the acquisition conditions onboard the carrier are recorded. These data, called metadata, are required to facilitate the further processing of RS data.

Metadata make it easier to search for RS data in the archives of data providers based on the parameters of interest. Such a search process typically includes metadata information for the identification of geographic coordinates, the acquisition time, calibration, etc. (ISO 19115-2). Thus, a painstaking and correct collection of metadata is indispensable for managing, storing, updating, searching, retrieving, using, and reusing RS data and for informing potential users. The simplest format used for metadata is the ASCII format (as applied for Landsat 5 data).

10.4.2 Housekeeping Data

Besides raw RS and metadata, housekeeping data are collected [67]. Distinguish between science housekeeping data (ancillary), engineering housekeeping data (ancillary), and spatial data (ancillary/derived):

Science housekeeping includes information about times (e.g., for the mission and the ephemeris), integration duration, status flags (indicating modes, problems, or events), and commands (e.g., histories, configurations such as the voltage offset, and data summaries regarding quality and signal strength). It is delivered by onboard sensors, the controller, or the command and data-handling computers.

Engineering housekeeping includes information about times, status flags, and commands, including goals, thresholds, configurations (for example, of deployables, tuning parameters of motors, switches, or sensors, or duty cycles), electrical states (of thermal systems, power supplies, etc.), and other physical and software monitors. These data are delivered by onboard sensors, controllers, and command and data-handling systems.

Spatial data are delivered to some extent by onboard systems. However, most of the spatial data are generated during the processing of L0 or L1 data. Examples of spatial data include positions, angles, distances (for example between the sensor, the Earth, and the Sun), orientations, status information (pointing, field of view, downlink, etc.), analysis flags (such as illumination, field of view, footprint geometry), and data on the target ID, locations, and physical parameters.

10.4.3 Additional Data

Additional data can be incorporated into the processing of RS data. In contrast to metadata (Sect. 10.4.1) and housekeeping data (Sect. 10.4.2), they are not linked to the acquisition or processing of the satellite data of interest. Additional data comprise data from (i) geoscientific sources, such as in situ data, modeling data, and value-added data, as well as (ii) other RS data that were, for example, acquired at a different point in time, at another location, and/or by another sensor or carrier. They can be used as spatial orientation, status information, and operative training data sets for the evaluation of algorithms or as training data sets for neural networks. During RS data processing, various preprocessing steps such as georeferencing [69], atmospheric correction [31], and the extraction of thematic information [70–72] can be optimized with the help of additional data.

Additional data can be used to spatially, temporally, and thematically define the object under investigation, to control search queries, and to further process RS data and extract information from that data.

10.4.4 In Situ Data

In situ data are the results of measurements and/or observations in the environment that are documented directly or in the immediate proximity of the phenomena or objects of interest in order to understand environmental relationships and/or processes [2]. Such data are required for the calibration of RS sensors and to set up, calibrate, validate, and update models that deal with RS data [73]. In situ data can be collected by scientists and citizens with or without deploying tools or by using autonomously operating instruments.

In situ data are always associated with errors and uncertainties [2]. They are introduced, for example, by the measurement itself (e.g., through measurement uncertainty and inherent limits on the accuracy of the measuring instrument). Furthermore, the operator or the measuring instrument can influence the data collected if the proximity of the operator or instrument affects, for example, the illumination conditions, the propagation of gaseous and liquid media, or the vegetation structure. Other error sources include a nonoptimal sampling strategy (e.g., unrepresentative sampling or repeated destructive sampling), not applying the measurement protocol strictly enough, and insufficient rigor of the operator.

An overview of sources that provide or link to freely available in situ datasets for various thematic fields is supplied in Table 10.9. The categories listed in Fig. 10.7 refer to parameters relating to (i) the atmosphere: aerosol optical thickness, atmospheric composition, clouds; (ii) agriculture: biophysical parameters, crop type, yield; (iii) the biosphere: biomass, biodiversity, vegetation structure; (iv) fire: bushfires, forest fires, intensity; (v) forests: biomass, growing stock volume, tree mortality; (vi) the cryosphere: ice-sheet soundings, sea ice; (vii) geophysical/terrain information: altimetry, gravity field, magnetic field, topography; (viii) the land surface: land cover, land use, land surface temperature; (ix) rivers/lakes: river discharge, salinity, water-leaving reflectance; (x) maritime information: salinity, surface roughness, sea ice; (xi) meteorology: temperature, water vapor, wind speed; (xii) snow: snow water equivalent, snow thickness; (xiii) soil: moisture, texture, types; (xiv) radiation: cosmic radiation, irradiance, global radiation; (xv) urban areas: urban heat islands, building structure; and (xvi) vegetation: biophysical parameters, chlorophyll fluorescence, and vegetation type.

10.5 Processing of Remote Sensing Data

The processing of RS data may include the following operational steps that depend on the degree of information aggregation (processing level): (i) data reception, transcription, data archiving, and quick-look processing within an operational environment of the ground segment [74, 75], (ii) preprocessing, which can comprise, e.g., geometric correction, atmospheric correction, and data fusion; (iii) thematic processing to obtain information (band combinations for deriving indices and classification results); and (iv) postprocessing to provide information for further applications (such as mapping, generation of GIS layers, and parameter tables) (Fig. 10.7). In the context of data acquisition by drones and airplanes, the first step (data reception, transcription, and data archiving) is often carried out during the first processing operations after the system has landed. The data processing can take place interactively, semi-automatically, or automatically.

The processing of RS data involves extracting qualitative and quantitative geoinformation as optimally as possible. This can be done interactively or in a semiautomated or automated manner.

Interactive processing has a long tradition and is usually performed for single images or small data series. Initially,

Table 10.9 Overview of selected in situ databases

Database name	Topic																Online resource
	Agriculture	Atmosphere	Biosphere	Cryosphere	Fire	Forest	Geophysics/terrain	Lakes/rivers	Land surface	Marine	Meteorology	Soil	Snow	Radiation	Urban areas	Vegetation	
Alberta Wildfire Historical Wildfire DB					X												wildfire.alberta.ca/resources/historical-data/historical-wildfire-database.aspx
ANGERS	X	X															opticleaf.ipgp.fr/index.php?page=database
AusCover TERN		X	X			X											www.auscover.org.au/dataset_categories/field-survey-datasets/
CEOS Cal/Val Portal	X	X	X			X		X		X							calvalportal.ceos.org/2
Collaborative Research Centre/Transregio 32	X	X	X				X		X			X				X	www.tr32db.uni-koeln.de/site/index.php
Copernicus In Situ Component	X	X	X	X	X	X		X	X	X	X	X	X			X	https://insitu.copernicus.eu/data-access
data.gov.uk	X	X				X	X	X	X	X	X		X		X		data.gov.uk/
EDI Data Portal	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	portal.lternet.edu/nis/home.jsp#
ENMAP Campaign Portal	X		X			X						X			X	X	www.enmap.org/flights.html
ESA Earth Observation Campaigns data	X	X	X	X	X	X	X		X	X		X	X		X	X	earth.esa.int/web/guest/campaigns
European Fluxes Database Cluster	X	X	X		X	X					X		X			X	gaia.agraria.unitus.it/home
Eurostat DB	X					X		X	X								ec.europa.eu/eurostat/data/database
FAO Soils Portal							X					X					www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/en/
Forest Observation System			X			X										X	www.forest-observation-system.net/
ForestPlots.NET			X			X										X	www.forestplots.net/
German Environment Agency	X	X	X		X	X		X	X	X	X	X		X	X	X	www.umweltbundesamt.de/en/data
GFZ Data Services	X	X	X	X		X	X	X	X		X	X		X	X	X	dataservices.gfz-potsdam.de/portal/

Table 10.9 (Continued)

Database name	Topic																Online resource
	Agriculture	Atmosphere	Biosphere	Cryosphere	Fire	Forest	Geophysics/terrain	Lakes/rivers	Land surface	Marine	Meteorology	Soil	Snow	Radiation	Urban areas	Vegetation	
Global Cryosphere Watch				X													globalcryospherewatch.org/data/data.html
GreenSeas										X							greenport.nersc.no/web/guest/database
GBOV of Copernicus Global Land Products	X		X		X	X			X			X		X	X	X	data.gbov.acri.fr/index.php?action=home
INSPIRE	X	X	X			X	X	X	X	X		X			X	X	inspire.ec.europa.eu/
International Soil Moisture Network									X		X	X	X				www.geo.tuwien.ac.at/insitu/data\protect_viewer/
LIDAR Online							X		X						X		www.LIDAR-online.com/
LOPEX	X	X															opticleaf.ipgp.fr/index.php?page=database
MODIS Land Team validation	X			X	X	X			X				X	X	X	X	landval.gsfc.nasa.gov/#
NASA's EOSDIS DAACs	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	earthdata.nasa.gov/about/daacs?_ga=2.188779627.1525515860.1547143706-1895070568.1547143706
NEON	X	X	X			X	X	X	X		X	X	X	X	X	X	data.neonscience.org/home
NOAA Digital Coast							X		X						X		www.coast.noaa.gov/dataviewer/#/
NOAA Earth System Research Laboratory		X									X			X			www.esrl.noaa.gov/gmd/dv/data/
OpenTopography			X			X	X	X	X						X		opentopography.org/
Pangaea	X	X	X	X		X	X	X	X	X		X				X	www.pangaea.de/
PSS-archi									X						X		www.pss-archi.eu/immeubles\protect_villes.html
RxCadre		X			X	X	X		X		X					X	www.firelab.org/project/rxcadre-project
TERENO Data Discovery Portal	X	X	X					X	X		X	X	X	X		X	www.tereno.net/ddp/
TRY-Plant Trait DB			X													X	www.try-db.org/TryWeb/Home.php
US Interagency Elevation Inventory							X	X	X								coast.noaa.gov/inventory/
USGS Earth Explorer							X		X	X							earthexplorer.usgs.gov/
WMO		X	X					X			X						www.wmo.int/cpdb/

10

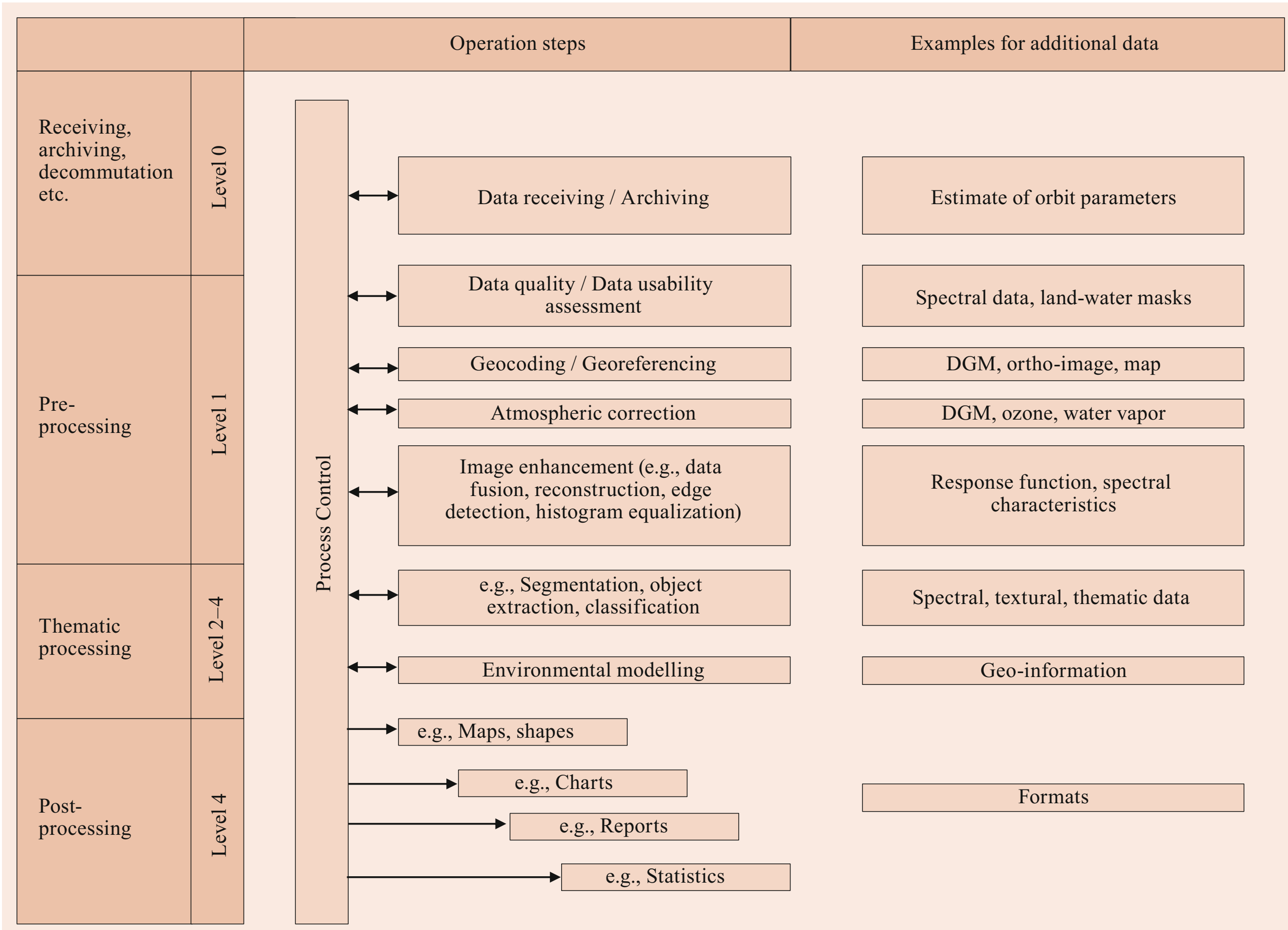


Fig. 10.7 Schematic of an operational processing chain for optical data, including the operational steps relating to the product aggregation of RS data (product level) and the housekeeping and additional data deployed

the evaluation was performed using optical interpretation devices such as stereoscopes and additional interpretation keys that were developed to understand the image content. Later, scientific and commercial image-processing systems for the computer-aided evaluation of data became increasingly popular [2]. Additionally, this processing mode is often used as pathfinder mode to identify, understand, and optimize the processing steps required to develop automated processing chains. The image characteristics that are considered in this mode are the context, color, area, shape, and texture of objects.

Due to the increasing supply of RS data, more and more effort is being made to optimize standard tasks in terms of costs, personnel, and time in order to ensure that the primary work—information extraction—is the main focus. Therefore, operations such as calibration, georeferencing, and atmospheric correction of, for instance, optical RS data are increasingly automated so that the data can be made available to the user at a higher processing level in real time or near real time (complex information products are made available < 180 min after data reception, [76, 77]).

The semiautomated processing mode is often used when visual monitoring of the data evaluation by an operator or interpreter is desired by the client, or when the intuitive and cognitive evaluation performed by an interpreter is not yet understood in detail or has not been implemented algorithmically.

Automated data processing is aimed at the processing of extensive datasets or time series. One of the first automatic processing chains was developed for the processing of Landsat data [78]. For example, the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS, [79]) was developed to produce standardized top-of-atmosphere (TOA) products from the Landsat 5 Thematic Mapper (TM) and the Landsat 7 Enhanced Thematic Mapper Plus (ETM+).

The development of corresponding processing chains is supported by an intention to provide information in an optimized and efficient manner, minimizing the subjective influence of manual processing in order to raise the product quality to a higher standardized level. Data processing can take place unsupervised (automatically) or supervised (interactively).

A scheme showing the different product levels in the context of an operational processing chain as well as the processing steps associated with those product levels for optical data is presented in Fig. 10.7.

10.5.1 Preprocessing

Besides data receipt, transcription, and archiving, preprocessing comprises image-processing procedures such as identifying regions of interest, data calibration, data usability assessment, geocorrection, atmospheric correction, and additional procedures to prepare the data for further processing, including (i) error correction algorithms for lost pixel values, lines, and image regions, (ii) histogram manipulation, (iii) noise reduction by filtering, (iv) pan sharpening, and (v) segmentation.

Data Quality Assessment

Obtaining a generally valid definition of the quality of RS data is problematic, since the data quality can be influenced by mission-related aspects (e.g., flight path disturbances), technical aspects (e.g., damage to the satellite, sensor disturbances or defects, and shading of the primary sensor by the satellite), subjective aspects, and environmental aspects (e.g., deterioration of the signal-to-noise ratio (SNR) due to the attenuation of the signal from the satellite to the ground station). Additionally, atmospheric phenomena such as haze, fog, and clouds can impact the data quality. Here, specific requirements of the intended application of the data are relevant. In atmosphere-oriented applications, atmospheric phenomena are the focus of investigation, while these phenomena are regarded as disturbing factors in, e.g., land- or water-oriented applications.

In the case of optical RS data, the data quality is usually assessed automatically or visually by an interpreter for a scene or individually for each quadrant of a scene and the result is stored in the metadata of the RS data [80]. A distinction is made between two measures, the cloud cover degree (the ratio of the number of cloud pixels to the total number of pixels of the evaluation unit, [80]) and the data usability (a combination of information about the cloud cover degree and the distribution of clouds within the evaluation unit, [81]). The data quality can be evaluated interactively according to a catalog of criteria [81] or by an automatic evaluation [80, 82].

Image Enhancement

An essential aspect of the preprocessing of RS data is image enhancement. The data from different sensors require different types of preprocessing depending on their characteristics. Documented image distortions for satellite data include:

- Due to a failure of the Scan Line Corrector (SLC) of the Landsat 7 Enhanced Thematic Mapper (ETM) sensor, the data has wedge-shaped gaps on either side of each scene. This data loss is about 22 % [83, 84].
- Memory effects in detectors, coherent noise, dropped lines, striping [84]
- Vertical stripes, such those observed in ASTER or TET-1 data, as a result of different radiometric sensitivities of the detectors in the sensor
- Missing measured values of individual pixels in the data
- The speckle effect in radar data.

In addition to image enhancement techniques that only consider individual pixels, algorithms that consider the neighborhood of a pixel, as captured by a static or moving kernel, are also used. Furthermore, methods such as fast Fourier transformation (FFT) are used to minimize or eliminate image interference [85].

Noise Removal

During digital image acquisition, the image can be contaminated by image noise. Noise is a random undesired variation in the image intensity in RS data that deteriorates the data quality [86]. There are three main types of noise [86, 87]:

- *Gaussian white noise*, which is randomly generated by electronic devices
- *Poisson noise*, which is mainly generated in the process of photoelectric conversion, where the influence of noise is more obvious in the case of weak light
- *Speckle noise*, which is a special feature of SAR imaging system images [48, 49].

Reducing image noise is an important aspect of quantitative data processing. The corresponding linear and nonlinear filters used include the mean filter, median filter, adaptive filter, Wiener filter, maximum and minimum filter, midpoint filter, and the alpha-trimmed mean filter [86, 88].

Band Correlation

One problem that can occur when recording RS data is misregistration between bands. This misregistration is one of the major sources of error when merging data from different data sources and quantitatively evaluating those data [89–91]. It can be caused by random movements of the RS carrier or by the parallax between the bands (different viewing angles), especially for mountainous terrain.

As a result of this misregistration, artifacts occur due to blurring between the bands. Therefore, a registration accuracy of less than one-fifth of a pixel should be aimed for in order to achieve a change detection error of less than 10% [90].

Georeferencing

Usually, georeferencing is the assignment of a geographical location to information [92]. The spatial evaluation of RS data using related geoinformation requires that the internal coordinate system of the RS data is matched with the three-dimensional coordinate values of a coordinate reference system such as latitude and longitude or Universal Transverse Mercator (UTM). Georeferencing is the process of linking features detected by RS with the corresponding objects and their coordinates in the real world [93].

During the acquisition of RS data, various geometric errors are introduced, and the effects of those errors on the RS data must be minimized. Sources of geometric distortions can be divided into systematic predictable and nonsystematic unpredictable sources [94–96]. In this respect, exact georeferencing aims to eliminate image distortions caused by, e.g., the curvature and rotation of the Earth, carrier movement, the finite scan rate and/or wide field of view of the sensors, rough terrain, central perspective recording, and/or incorrect orientation of the imaging system. Surveys of algorithms used in practice and/or other applications are given in, e.g., [3, 95, 97–100].

Two different approaches are used to correct geometric distortions. The first approach, which is employed to correct systematic distortions, involves mathematically modeling all distortion sources (in nature and magnitude), and thus requires accurate and comprehensive knowledge of all possible sources of interference during the scene recording time. The models created during the modeling are then used to create correction formulas that are applied to the RS data.

The second approach was developed to correct random distortions as well as unknown systematic distortions, and requires a rectification base (e.g., a map or orthophoto) so that spatial relations between the image pixels of identifiable objects and the corresponding geographical coordinates of the same objects in the rectification base can be identified, which in turn allows mathematical relationships and thus mapping polynomials to be formulated [94]. This method requires the selection of a number of clearly recognizable pixels (ground control points: GCPs) in the distorted image (i.e., by row and column) and their assignment to either (i) their actual positions in ground coordinates e.g., latitude, longitude; GNSS or dGNSS survey, digitized from a map, for image-to-map registration or (ii) a previously georeferenced image (equalization base) for image-to-image registration, whose pixels are already linked to the coordinates of a geographic coordinate system by mathematical transformation [94].

After registration, resampling is performed to identify the points in the distorted image that correspond to the positions of the pixel grid previously defined on the map [94].

Reference [94] points out that the assignment of the pixel values in the distorted image to the pixel positions in the rectified image is not always clear; in some cases it is necessary to decide which pixel brightness value should be selected for placement in the new grid. Three common resampling methods are used for this processing step: nearest neighbor, bilinear interpolation, and cubic convolution [94].

The efficient use of RS data in an operational manner may require orthorectification. If it is performed based only on the ephemeris and additional attitude data measured onboard the satellite, geocoded data of varying accuracy may result [101]. However, if the required absolute accuracy is in the subpixel range, more precise georeferencing is required. Therefore, parameters such as the required metadata, the absolute accuracy, the computing time, the potential for optimization and automation, and the usability and functionality of an algorithm must be considered [101]. In order to enable the automation of the georeferencing of RS data, the raw data from various satellites (such as GeoEye, DigitalGlobe) are delivered with so-called rational polynomial coefficients (RPCs), which are stored in the supplied metadata. RPCs enable higher systematic modeling of the georeferencing because they contain information about the viewing geometry, thus theoretically enabling orthocorrection through the inclusion of a digital elevation model (DEM) and the use of 3-D calculation rules, such as those for calculating terrain, forest, and building heights.

Calibration and Radiometric Correction

Since the energy recorded by a sensor element can differ from the energy actually reflected or emitted by the measured ground element, a radiometric examination of the measured values is necessary. These radiometric differences are determined by the intervening atmosphere, the Sun–sensor geometry (the azimuth and elevation of the Sun; the position of the sensor in relation to the object and the Sun/illumination), and errors in the sensor itself.

Thus, radiometric correction is performed to (i) calibrate the sensor elements and thus avoid radiometric errors in pixel values and (ii) minimize radiometric errors or distortions in pixel values introduced when the data are recorded, so that the quality and therefore the interpretability of the recorded data are improved.

Additionally, a sensor records the measured intensity of the electromagnetic radiation as a digital number DN [68]. These values can be converted to real physical values (known as the at-sensor radiance) using the satellite-band-specific rescaling factors provided by data providers in the metafiles [68]. By using additional information provided in the metadata, such as day of year and the local Sun elevation angle, physical parameters such as the Earth–Sun distance and the solar–zenith angle can be computed and then used (in combination with the exoatmospheric irradiance) to compute the so-called top-of-atmosphere (TOA) reflectance.

After applying an atmospheric correction, which requires information on the atmospheric conditions at the time that the RS data were recorded, the surface reflectance can be computed.

Atmospheric Correction

The integration of RS data into environmental modeling, image mosaicing, the detection of changes at the Earth's surface, and the delivery of standardized data sets requires the derivation of physically corrected input parameters. Therefore, atmospheric correction is a prerequisite for the derivation of quantitative parameters from RS data [102].

Electromagnetic energy passes through the atmosphere twice: from the Sun to the Earth and back to the sensor. Along its path through the atmosphere, absorption and scattering processes modify the electromagnetic radiation. Whereas absorption reduces the intensity of electromagnetic radiation, leading to a haziness effect, the scattering redirects electromagnetic energy in the atmosphere, causing an adjacent effect where neighboring pixels are shared. These two wavelength-dependent processes affect the quality of an image and are the main reasons for performing atmospheric correction [103].

Optical RS is especially influenced by clouds and haze. Atmospheric correction removes the effects of atmospheric scattering and absorption to obtain the surface reflectance [31].

Dark object subtraction, radiative transfer models (LOWTRAN: LOW-resolution propagation model for predicting atmospheric TRANsmittance; MODTRAN: MODerate resolution atmospheric TRANsmission), and atmospheric modeling are common techniques that are used to correct for atmosphere disturbances. Examples include the Dark Object Subtraction technique [104], the dense dark vegetation method [105], the Simplified Method for Atmospheric Correction [106], and the Second Simulation of the Satellite Signal in the Solar Spectrum (6S) atmospheric model.

Dark Object Subtraction (DOS) is the simplest atmospheric correction method for RS data. This approach is based on dark objects such as lakes, coniferous forests, and dark asphalt (which have surface reflections of zero or almost zero), their correct identification [107], and a spatial distribution that permits subsequent correction of the data [108]. Considering these preconditions, the DOS method assumes that the minimum DN value of a histogram (if result of a dark object) is subtracted from the DN values of the pixels of the entire scene because this result is the atmosphere [109]. The amount of additional data required to perform this procedure is small.

Precise atmospheric correction requires the solution of atmospheric radiation transfer equations. Therefore, atmospheric modeling is the most sophisticated correction method. These modeling approaches are ideal when scene-specific atmospheric data (aerosol content, etc.) are available.

Examples of propagation models for electromagnetic radiation include the Simplified Method for the Atmospheric Correction (SMAC; [106]), LOWTRAN [110, 111], MODTRAN [112, 113], the EXact Atmospheric Correction Technique (EXACT, [114]), and the 6S code [36].

However, given that actual atmospheric information is rarely available, the correction requires special expertise. Additionally, atmospheric correction is time-consuming, although the correction procedure has been accelerated by calculating look-up tables for varying wavelengths, solar zenith angles, satellite viewing angles, and optical thicknesses of aerosols [32, 115, 116]. These tables are used in, e.g., interactive correction methods (e.g., ATCOR, [103, 117, 118]).

Interactive atmospheric correction is applicable for a small series of data sets. However, in order to correct a large number of data sets, an automatic correction procedure (e.g., ClearView: [31]; automated version of ATCOR: [103]) that performs standardized data correction in an operational environment and minimizes subjective processing errors is required [31, 119]. In particular, in parallel with RS data acquisition, the ClearView procedure is designed to obtain actual atmospheric and additional data collected by different satellites, e.g., an elevation model and data on water vapor, ozone, aerosol quantity and type, and clouds/shadows [31]. The state of the atmosphere above a study area on the Earth's surface at the exact time that the satellite flies over is determined by various methods (e.g., physicochemical models in combination with assimilation techniques). This is necessary because not all of the data required for the correction can be collected at the same time of day. The large atmospheric variability, which leads to considerable errors, can be compensated for by evaluating measurements in different spectral channels.

The atmospheric correction of optical RS data and time series of data is a prerequisite for the quantitative evaluation of such data, and ensures data quality and consistency. Therefore, [120–123] propose that the aerosol optical thickness (AOT) calculated from RS observations should be compared with the AOT calculated using the Aerosol Robotic Network (AERONET). Reference [121] also propose a cross-comparison methodology for RS data that takes into account the surface anisotropy by including the bidirectional reflectance distribution function (BRDF) and allows spectral adaptation of RS data (Terra and Aqua-MODIS, Landsat 5/TM, Landsat 7/ETM+) using an artificial neural network trained with the radiation transmission model PROSAIL.

Analogous to the ClearView processor [31], when using the Sen2Cor processor framework, it is necessary to designate a classification for the processing, such as for example, the cirrus clouds correction. In addition to these data, [31, 122, 123] propose the use of additional data such as digital elevation models. The use of ECMWF data can improve the quality of processing [121].

Data Fusion

Data fusion aims at combining complementary data (e.g., RS missions, sensors, and data of spatial and temporal resolution originating from different data sources) in order to minimize the disadvantages of the source data in the fused data set and to increase data quality [124]. The expected result is a fused data set with higher levels of consistency, precision, and usefulness than can be achieved using only the individual data sources.

In the context of RS, different fusion techniques can be distinguished according to the level of information aggregation, such as pixel level (measurements), feature level (identified and extracted objects), and decision level (rule-based and decision-based analysis), or according to the source and type of input data (single sensor–temporal; multisensor–temporal; single sensor–spatial; multisensor–spatial; single data–multisensor; fusion with additional data or topographic maps; [125]).

The intended improvements in the processed data are, for example, image sharpening, geometry correction, stereoviewing, the detection of features that are not visible in the individual source data sets, the addition of missing information, and the replacement of defective data.

One important data fusion method is called pan-sharpening. In this process, multispectral data of low spatial resolution are combined with panchromatic data (or monochromatic data) of higher spatial resolution in order to obtain multispectral data of higher spatial resolution.

A basic principle of data fusion is color model transformations (such models include hue–saturation–value: HSV, [126]; hue–lightness–saturation: HLS, [127]; and Intensity-Hue-Saturation: IHS, [128]). Other methods such as arithmetic channel combinations, principal component analysis, wavelet transformation, and regression variable substitution evaluate the input data statistically or arithmetically and transform them into the intended result of fusion.

The quality of the fused image data depends, among other things, on the preprocessing of the input data and the selected image fusion method. The more that the physical characteristics of the input data are considered during data fusion, the easier it is to select suitable preprocessing and fusion methods.

The suitability of a fusion technique depends on the intended task or application, the sensor characteristics, the signal-to-noise ratio of the sensor, and the data availability, among other factors. Further temporal aspects such as the time that the data were recorded, the temporal relationship between multitemporal recordings, and changes in the object of interest between data recordings are important factors [129].

High-quality georeferencing and accurate coregistration of the input data are prerequisites for successful data fusion.

The comparison of different fusion algorithms based on the fidelity of the images they yield is a complex task [130]. Different measures for assessing the quality of the fused data have been adopted (e.g., the correlation coefficient (CC), the difference in variance (DIV), the standard deviation (SD), the root mean square error (RMSE), and the spectral angle mapper (SAM); [131]) or developed (e.g., the Universal Image Quality Index (UIQI), [132]; the Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS), [131]). These test algorithms focus on image properties as indicators of the preservation of spectral characteristics, the increase in spatial resolution, the level of detail for textural criteria, and the minimization of information loss.

Data fusion aims at the creation of a new data set that is more suitable for processing by humans or computers in terms of deriving the maximum information content [133], feature extraction, and object identification. Accordingly, data fusion is one of the most common steps in information acquisition and the visualization of complex phenomena and correlations in geoinformatics.

10.5.2 Thematic Processing for Information Extraction

Thematic information that the user is interested in, such as the physical and chemical properties of an object, are represented only indirectly in RS data (Sect. 10.4). Therefore, image features such as (multi)spectral signatures, transformed parameters (e.g., spectral indices), texture parameters (e.g., homogeneity, heterogeneity, shape, and edges), segments and their neighborhood relationships, and object and time information are deployed to extract information [5, 134, 135]. Background information regarding typical extraction methods such as edge extraction, line tracking, texture feature classification, shape feature classification systems, look-up tables, decision trees, the minimum-distance classifier, and the maximum-likelihood classifier is provided, for instance, in [134–137]. Information extraction methods are selected application-specifically to ensure optimal results. The information extraction itself requires a number of processing steps [138] to systematize properties, objects, or phenomena into categories with comparable characteristics (such as properties or behavior). The overall goal of the delimitation in the feature space is a simplification to clarify complex facts and increase their understandability. The extraction of the required information from RS data can be carried out by statistical (*Classification*), physically-based (*Physically-Based Methods*), and machine learning (*Machine Learning Methods*) methods.

Classification

In principle, classification aims at the intended assignment of the data elements of concrete observations to data classes. The elements can be differentiated from the elements of other classes by considering decision criteria [139]. Image classification is a problem-induced operation focusing on the qualitative or quantitative thematic evaluation of RS data in order to deliver interpretation results for many environmental and socioeconomic applications [139, 140]). The classification process comprises the segmentation of image regions according to rules and factors for similarity and the labeling of identified image regions.

Factors that disturb the classification performance include atmospheric factors (e.g., absorption and scattering processes) and terrain-related factors (e.g., ditches, slopes, differences in lighting), all of which must be eliminated or at least minimized [31], as they can impair the result.

The taxonomy of classification procedures is: supervised classification approaches, unsupervised classification approaches, parametric classifiers, nonparametric classifiers, per-pixel classifiers, object-oriented classifiers, per-field classifiers, hard classifiers, fuzzy classifiers, spectral classification, contextual classifiers, and spectral-contextual classifiers [140, 141]. According to [140], the preparation, execution, and postprocessing of a classification of RS data can include the following detailed steps: the determination of a suitable classification system (e.g., methodology) and class boundaries, preprocessing of the RS data, feature extraction, selection of the training data, and accuracy evaluation.

Although classification often forms part of the main processing step in image processing, classification is also performed in preprocessing steps such as data usability analysis [142], geocorrection [31], and data fusion [143] to assign properties such as the optical thickness, bidirectional reflectance distribution, emission values, or geometric properties (such as the center of gravity, area, perimeter, and so on) to areas for further use.

In addition to conventional classification algorithms such as the decision tree, minimum distance, maximum likelihood, and cuboid classifiers [144–146], artificial classifier methods such as neural networks [147], the fuzzy-set classifier, and expert systems such as the cross-validation classifier [148], the random forest classifier [149], or the support vector classifier [149, 150] that are collectively termed *deep learning methods* are increasingly being used to classify RS data.

Additionally, advanced classification approaches and techniques for improving classification accuracy have been developed, e.g., evident reasoning approaches [151], the object-based, fuzzy-based classification technique [152], and per-pixel classification combined with contextual image information [153].

An essential aspect of the classification is the analysis of the achieved classification accuracy. Ground truth data (or in situ data) are needed to evaluate the quality of a classification. However, it must be considered, that beside possible unsystematic and systematic measurement errors of in situ data, there exist differences in spatial and temporal dimensions in between in situ measurements and remote sensing measurements. Additionally, it should also always be asked whether the same variable is measured that makes it possible to validate remote sensing data.

In RS, the error or confusion matrix [154, 155] has become an established method. This involves constructing a square matrix with the same class listed in the rows and columns.

A standard procedure used to verify the quality of classification results is based on the evaluation of a confusion matrix [154, 155]). There are various quality measures, such as the overall accuracy (the proportion of correctly classified pixels among the total number of pixels in the input data), the user accuracy (the percentage of pixels of a class in the product that are correctly classified), the manufacturer accuracy (the percentage of pixels of a class in the reference data set of the product that are correctly classified), and the kappa coefficient (which verifies the conformity of the classification product with the reference classification, taking into account the commission error and omission error).

Physically-Based Methods

Physically-based methods aim at the retrieval of object properties from RS data through the transformation of the acquired radiation into variables of interest based on physical laws [66]. The physical laws are deployed to describe the interactions of photons with the media that they pass through on their way from the emitting source to the receiver [66, 156]. The presumed framework conditions, physical laws, and knowledge about the interoperation of both are summarized in radiative transfer models (RTMs, [66, 157]).

RTMs are designed to simulate the propagation of radiation in a medium [157] such as the air, aerosols, clouds, vegetation canopies, forests, plant leaves, soils, water bodies, and snow [73, 158]. For this purpose, the medium is divided into small elements of thickness d*s* with certain properties [43]. Inhomogeneities in the medium are described by variations in the properties of the elements constituting the medium. In order to determine the intensity of the radiation that leaves an element in a certain direction, an equation that considers (i) the direction, wavelength range, and intensity of the radiation that enters the element [27], (ii) the reflection and scattering in the element, and (iii) the attenuation and generation of radiation within the element [43] is required. By combining the equations of all the elements in the medium into one radiative transfer equation, a formal description of the propagation of radiation in the medium is obtained [27]. A deterministic description of the influences to which the radiation is exposed in several media can be attained by coupling various compatible RTMs [157, 159].

RTMs can be operated in two modes. In the forward mode, properties of the medium and the spectral, angular, and spatial characteristics of the sensor are known and used as inputs for RTMs that can predict the corresponding spectra [66]. Modifications to the properties of the medium cause the interactions of the photons with the medium to change, which in turn affects the characteristics of the simulated spectra [66, 157]. In the inverse mode, spectral data as well as the spectral, angular, and spatial characteristics of the sensor are required as input. The RTMs are deployed to interpret the spectral observations in terms of the state variables that characterize the medium [66, 156]. This means that the values of the state variables that generate the measured spectrum need to be estimated [160]. To ascertain these estimated state variables, a cost function is set up.

Cost functions aim at minimizing the distance between either simulated and measured spectra (the radiometric data-driven approach) or the state variables derived in the inverse mode from the measured spectra and those of a reference data set (the state-variable-driven approach, [66]). Both approaches deliver a proper solution if and only if there is a solution that is unique and depends continuously on the data [161]. In general, the inverse problem breaks with at least one of these criteria, and is therefore not well posed (or is *ill posed*, [160]) for the following reasons: (i) the amount of independent information contained in the spectral measurements is (often) insufficient to uniquely determine the unknown state variables [66]; (ii) the solution is not unique, since several combinations of state variables result in simulated spectra that deviate to a similar extent from the measured spectra; and (iii) deviations from the true spectrum may arise from uncertainties in the measurements and the model [160].

In order to derive a stable and reliable solution, the problem needs to be regularized. One option to constrain the problem is the integration of prior information [66]. Prior information can be provided, for example, in the form of additional data acquired in situ or by other terrestrial, airborne, or spaceborne RS systems [160]. Furthermore, knowledge of the distribution and behavior of state variables originating from past or comparable experiments, experts, or bibliographies can be incorporated as prior information. The regulative power of prior information increases as definitions become stricter. Strict definitions are enabled if the domains for which they are valid are divided in subdomains [66]. In general, the complexity of the general radiative transfer equation can be reduced by inferring separate RTMs for designated wavelength ranges (e.g., VIS and NIR, TIR, or MW), since different physical phenomena dominate the interactions between photons and the medium in various wavelength ranges (Fig. 10.1e, [13]). When accompanied by a precondition that only sensitive, independent, and relevant state variables should be used, this approach leads to improved solvability [43]. Furthermore, the consideration of spectral information that was acquired previously or in the neighborhood of the location allows for the incorporation of constraints describing the temporal or spatial evolution of the state variables, respectively [66].

Machine Learning Methods

Artificial intelligence (AI) is regarded as one of the key technologies in future societal and technological development. Although this technology is now being introduced into many areas of technological development, its history dates back to the 1930s [162].

AI is increasingly being applied to the evaluation of image understanding and therefore also to RS data. AI is a subsection of computer science that deals with the automation of intelligent behavior and machine learning [162, 163]. AI systems can be divided into rule-based systems, classical machine learning systems, and representation learning systems, which can in turn be subdivided into feature-based machine learning and deep learning (DL) systems. AI deals with the development of instructions for independently and intelligently solving a problem or a class of problems.

Machine learning includes procedures that serve the artificial generation of knowledge in a learning phase utilizing example data. The artificial system recognizes patterns and laws in the training data through generalization and abstraction. Machine learning technologies use DL techniques, which are particularly suitable for processing large amounts of data.

Big data are input into neural networks in order to train the system to classify this data accurately. After a successful learning phase, the system can assess unknown data (learning transfer) or overfit the unknown data [162–164].

In addition to classical statistical methods, machine learning methods are increasingly being deployed in RS evaluation in order to evaluate big data quickly and efficiently under, e.g., relevant environmental aspects. In this context, attempts are made to reproduce human decision structures in automatic computer models such that they can learn and then apply their learned knowledge to an ambiguous environment, enabling intelligent problem solving. DL algorithms are hierarchically structured and use neural networks; with increasing complexity and abstraction, such algorithms are able to process large big data [162, 164].

DL is now used in almost all fields of RS image processing and analysis, such as image fusion, image registration, scene classification, object recognition, land use and land cover classification, segmentation, and object-based image analysis.In the context of remote sensing data preprocessing, Stacked AutoEncoder and Convolutional Neural Networks have been used as successful DL models, with Recurrent Neural Networks and Generative Adversarial Networks also expected to be used in the near future [165]. In contrast, the use of DL methods in image registration is considered problematic, which is attributed to the lack of public training datasets [165]. For scene classification, object recognition, semantic segmentation and LULC classification, supervised DL models emerge based on large amounts of training data, while unsupervised DL models are considered successful, limitations in training data need to be overcome [165].

Machine learning methods and procedures can be used to solve optimization problems. DL is an essential part of machine learning and is based on the use of neural networks and big data. Neural networks represent a cascade of multiple layers of nonlinear processors for feature extraction and transformation. The results from a layer represent the input for the next layer.

The development of machine learning methods is driven by the aim to improve computer algorithms by integrating cognitive experience. Thus, the DL technique enables a computer model to learn how to perform classification tasks directly using sampled data (e.g., images and text). The model is trained on extensive classified data sets using a neural network architecture with many levels that represent different levels of abstraction. The learning process can be supervised learning or unsupervised learning. Corresponding learning algorithms are, e.g., decision tree learning, artificial neural networks, evaluating hypotheses, Bayesian learning, computational learning theory, combined inductive and analytical learning, and reinforcement learning [166]. The main issue is the learning process and the methods used to train and condition the networks. A largely comprehensive list of 64 machine learning methods (e.g., BN – Bayesian network, DNN – deep neural network, FCM – fuzzy C means, HCA – hierarchical clustering analysis, NN – neural network, PCA – principal component analysis, RandNN – random neural network, RF – random forest, RL – reinforcement learning, SVM – support vector machine, SVR – support vector regression) is given by [167]. These methods have been successfully used, for example, to classify RS data [168–171].

10.5.3 Postprocessing

Postprocessing aims to provide information derived from the RS data to the user. The information must be provided in a form that the user expects, i.e., that the user can understand and integrate into its processing, process sequences, and production processes. Accordingly, postprocessing includes all processing operations, (e.g., data improvement, change recognition, format transformation) that are suitable for preparing the data in an application-oriented and timely manner so that the user can utilize the information optimally and, ideally, negate the need for further processing steps.

Data are made available to the interested user in the form of images, maps, numbers, charts, and reports in a wide variety of data formats, e.g., pixel, vector, and text formats or combinations thereof.

The most common ways to present information are as an image, a sequence of images, or a map in pixel form. The postprocessing of thematically evaluated RS data can take place at different processing levels. Digital image postprocessing routines include the enhancement of images in order to fill data gaps and to minimize errors, improving data homogenization. These computerized process routines increase the image scene quality, thus leading to improved image interpretation.

In addition to the provision of information in pixel form, the information may also be requested in the form of vectors (e.g., in order to integrate it into GISs by using additional data for further evaluations). In this case, the evaluation results from the thematic processing are subjected to a further process step that is also referred to as vectorization [172]. The results are transformed from the pixel format into vectors according to previously defined rules.

In various cases, information is derived from the RS data in the form of reports, tables, and charts, which are then used, for example, in statistical applications for market analysis and by economic or political decision makers.

10.6 Applications

Applications of RS are diverse in type, quality, and quantity. The market drivers for the development of a RS market are (i) the availability of high-resolution RS data as a result of new satellites, sensors, and missions, (ii) easier user access to RS data and products through the further development of the Internet, (iii) lower data costs, (iv) the development of the geographic information system (GIS) sector, and (v) the development of new evaluation methods.

With the implementation of the Copernicus project (formerly GMES – Global Monitoring of Environment and Security), a joint initiative of the European Union and the European Space Agency (ESA), concrete steps were taken to develop a European geoinformation market by strengthening the space and ground segments as well as the use of RS data.

Other factors that have promoted the development of an international geoinformation market include (i) the increase in the number of nations operating RS missions, (ii) the entry of commercial market participants into, e.g., the space segment and the distribution segment, (iii) the increase in commercial companies that are marketing information products or services based on them, (iv) the rapid development of the telecommunications and advertising market, (v) the rapid development of computer technology, software technology, and Internet technology, (vi) the free access to the data pool of the Landsat program granted by the United States Geological Survey (USGS), and (vii) the free access to Sentinel mission data granted by ESA.

As described in the Introduction, RS data can be used in different applications and with different degrees of information aggregation. Therefore, the applications described below are only a representative selection of the applications that use RS data. In Table 10.10, the spatial scales considered in various geographic research areas are compared with map scales and geometric resolutions of objects in the landscape and with the pixel resolution of satellite data.

10.6.1 Remote Sensing for Weather and Climate Research

RS data are recorded for aspects of terrain climatology, environmental protection, and technical climatology using, for example, airborne radiometers. In addition, a number of im-

Table 10.10 Comparative analysis of different spatial scales considered in various geoscientific fields with the spatial pixel resolution of operational satellite data (adapted from [173]). Information to this aspect can be found in, e.g., [174]

Scale 1 : X (m)		Dimension										Spatial Resolution			
		Environmental Systems							Manmade Systems			Satellite		Map	
		Geosphere	Climatology		Hydrology	Pedology	Vegetation Geography	Ecology	Environmental programs	Administrative Level		Pixel Resolution		Object	Lineament
10^6	Small-Scale Map		Macro-Climate	Global Climate			Vegetation belt		COPERNICUS GRID CORINE	Community of States			21	4000 km²	1200 km
10^5		Geo-sphere		Zone Climate	Drainage Basins 1th-order	Soil Zone	Vegetation zone	Biogeo-sphere	LTER TERENO					400 km²	120 km
10^4				Main Type of Climate	Drainage Basins 2nd-order	Soil Region; Soil Province	Vegetation formation	Eco-System complex		State; Federal State		Very low	20 19	40 km²	12 km
10^3	Middle-Scale Map	Region	Meso-Climate	Landscape Climate		Soil society	Vegetation class		National Parks Program MAB	District				4 km²	1.2 km
		Core			Drainage Basins 3rd-order	Soil land-scape	Vegetation association	Eco-system		Community		Low	12 11 9 17 18		
10^2	Large-Scale Map					Pedo-complex	Vegetation socienty	Bio-cenosis	Long-term observation Areas			Medium	8 10 14 15 16	400 m²	120 m
10^1				Climate of a Site				Popula-tion		Neighbor-hood		High	5 7	40 m²	12 m
10^0		Top	Micro-Climate	Micro-Climate	Hydro-top	Pedo-top	Physio-top		Environmental measurement program	Plot		Very high	4 6	4 m²	1.2 m
10^{-1}				Interfacial Climate	Lysimeter	Pedon		Indivi-duum				Ultra high	2 3 13	0.4 m²	0.12 m
10^{-2}													1	0.04 m²	0.01 m

Environmental Observation Programs:
COPERNICUS – Global Environmental Monitoring System
GRID – Global Resource Information Database
CORINE – Coordination de l' Information sur l' Environment
MAB – Man and Biosphere
TERENO – Terrestrial Environmental Observation
LTER – Long Term Ecological Research

Ground Resolution:

Very low	(> 250 m)
Low	(50 – 250 m)
Medium	(10 – 50 m)
High	(4 – 10 m)
Very high	(1 – 4 m)
Ultra high	(< 1 m)

Satellite systems:

1 WorldView-4 / Pan	8 SPOT 1-4 / PAN	15 OCM-2 / LAC
2 WorldView-4 / MS	9 SPOT 1-4 / MS	16 Sentinel 3 / SLSTR
3 QuickBird / Pan	10 Landsat 7,8 / PAN	17 NOAA / AVHRR
4 QuickBird / MS	11 Landsat 7,8 / MS	18 MSG / HRV
5 RapidEye / MS	12 Landsat 7,8 / Thermal	19 MSG / SEVIRI
6 SPOT 6 / PAN	13 TerraSAR-X SpotLight	20 SMOS / Radar
7 SPOT 6 / MS	14 TerraSAR-X ScanSAR	21 SMOS / Radiometer

portant conclusions can be obtained regarding, for instance, air pollution and plant growth (intact or diseased plants, drought stress, diseases) or the state of agricultural development by analyzing RS data. In many cases, RS is used to determine the temperature distribution on the soil surface.

In addition to airborne RS, satellites and groups of satellites are used to obtain climate and weather information products. TIROS 1, the first weather satellite, was launched on April 1, 1960 [39, 175]). The satellites used for RS applications can be divided into two groups. Group 1 comprise geostationary satellites (see Sect. 10.3.4, *Space Segment* on the space segment) in 36 000 km orbits. Group 2 consists of polar orbiting satellites with orbital altitudes of between 100 and 2000 km [39].

An essential field of application of satellite RS is operational precipitation extraction. The evaluation methods developed for this purpose analyze infrared data [176, 177] and, increasingly, microwave data in order to collect precipitation events. In addition, satellite RS is used to obtain predictions of tropical cyclones [178], determine wind vector fields from GOES data [179], and determine cloud properties, aerosol properties, perceptible water, and profiles of temperature and humidity from MODIS data [180], and precipitation determination in regions with a sparse ground-based rain gauges network [181]. Environmental parameters such as cloud, aerosol, water vapor, and surface properties are also measured using airborne scanning spectrometers.

In addition to the image data deployed to create cloud images and time series [182–185], the water content, air pressure, and temperature at different altitudes of the atmosphere [175] can be determined from the radiation data in certain wavelength ranges. Wind direction and speed can be estimated from the movement of a selected cloud.

Besides stationary networks of rain gauges, stationary weather radar stations enable RS of rainfall from the ground, for example [186]. These systems transmit impulses in the frequency range 3–30 GHz towards the sky. The transmitted signals are partially reflected by rain or precipitation, and the distance to that precipitation can be derived from the transit time [187]. Output data from a dual polarization radar include the reflectivity factor (Z), Doppler velocity (V), and spectrum width (W; [187]).

The main advantages of a weather radar are its high spatial resolution and temporal resolution (5 min). In order to derive a simplified Z–R relationship, i.e., the relationship of the measured radar reflectivity factor Z (mm^6 m^3) to the rain rate R_{Int} (mm/h) [188–190], the regression parameters a and b must be derived, which are strongly location dependent and therefore show a strong spatiotemporal variability [190].

$$\int R = aZ^b \qquad \text{(mm/h)} \tag{10.22}$$

It should be noted that the reflectivity factor and the rain rate both depend on the drop size distribution.

10.6.2 Maritime Remote Sensing

Continuous observations of the oceans and ice are important in many ways. First and foremost, they provide data that are useful for (i) ensuring the safety and security of ocean shipping and offshore activities, (ii) monitoring infrastructure (e.g., protective structures, wind turbines, and port systems), and (iii) monitoring environmental phenomena and their dynamics (e.g., wind conditions, wave height, direction, and speed, and temperature). Detailed overviews of the utilization of RS for maritime applications are provided by [191, 192].

The monitoring of ice areas and their dynamics—whereby parameters such as the structure, thickness, age, and movement of the ice as well as the ice classes present are recorded—contributes to the safety of shipping. In addition, sea ice from the polar regions is an essential component in the circulation of the global climate system due to exchange processes between the atmosphere and oceans. RS data are thus essential for investigations of past, present, and future climate conditions and processes. Monitoring and realistic quantification of the size, quantity and time dynamics of the sea ice are therefore an important prerequisite for the development of realistic climate models.

Especially in the last few years, maritime applications of RS have played an important role in environmental monitoring. For instance, RS has been used to monitor overfishing of the oceans [193], oil spill pollution [194], and marine debris pollution (MDP, [195]), which have increasingly moved into the focus of general public interest. Tools such as webcams and digital cameras have enabled these problems to be monitored using RS (e.g., in the case of MDP, [196, 197]). Earth observation technology permits the assessment of various physical, biological, and ecological parameters of water bodies at regional and global scales.

Due to the fact that various processes are better understood today, references to RS can be established in order to develop appropriate monitoring systems.

Examples of important research fields involving the operative and automated application of RS are the monitoring and assessment of the ocean state. In this context, relevant parameters are the significant wave height, peak wave direction, and wavelength among ocean surface characteristics. A further aspect is wind field retrieval (the wind direction is discerned from wind streaks imaged on the ocean surface or the wind speed is determined from the ocean surface backscatter). Other applications that use optical, thermal, and/or SAR data include oil spill monitoring (e.g., locating oil spills and distinguishing between oil spills and lookalikes), ship detection and identification (e.g., locating ships, classifying ships in terms of length, width, etc., and the fusion of ship locations with time-correlated automatic identification system (AIS) data), topography/bathymetry monitoring, monitoring the dynamics of land masking, and iceberg detection (locating icebergs, detecting ice concentrations).

10.6.3 Remote Sensing of Inland Waters

In addition to oceans and seas, which can be considered very good targets for RS due to their size, inland waters are also of interest for qualitative and quantitative monitoring due to the freshwater resources stored in them and their ecological and economic significance. Inland waters include standing waters, e.g., lakes, ponds, fens, and raised bogs, as well as flowing waters, e.g., springs, streams, and rivers [198]. A distinction can also be made between these natural waters and artificial waters, e.g., dams and canals.

While area objects such as lakes are relatively good monitoring objects, rivers can be difficult target objects, as they are elongated (meandering or elongated) objects with comparatively small widths.

The spectral characteristics of water, suspended sediments, floating matter, and pollutants, which are important for assessing, e.g., the hydrological, biological, and chemical characteristics of the water, are essential factors in the monitoring and assessment of water quality. Thus, the possibility of obtaining RS-based information depends on the pixel resolution of the sensor used and the shape of the water body of interest. The water parameters of interest include water quality indicators, the total suspended solids (TSS), chlorophyll a concentration, turbidity, salinity, total phosphorus (TP), Secchi disk depth (SDD), temperature, pH value, and dissolved organic carbon (DOC, [199]).

In the case of standing waters, parameters such as the water transparency, chlorophyll a, and total phosphorus concentration [200], water quality [201, 202], total suspended matter (TSM, [203]), and temperature [204] have been investigated. In the context of flowing waters, parameters such as the chlorophyll a and turbidity [205], nutrient and suspended sediment loads [206], dissolved organic carbon, and turbidity [207] have been investigated.

In recent years in particular, RS has also been used to investigate fens and raised bogs, which have increasingly attracted the interest of scientists and the public due to their importance in the climate debate. References [208–210] have presented work on this topic.

10.6.4 Remote Sensing for Geomorphology and Geology

The quantitative and qualitative evaluation of the landform ensemble was initially carried out analogously using stereoscopic aerial photographs. Stereoscopic images were assumed to represent spatial models of the recorded sections of the Earth's surface [211]. Evaluating the geomorphologic treasure of forms of a landscape in stereoscopic images allows for conclusions to be drawn about the geology of and morphological events for the landscape. The interpretation of stereoscopic aerial photography is based on the detection of height differences, slope inclinations, river gradients, the thicknesses of geological layers, and their outcropping [211].

With the use of digital image processing methods with respect to RS, the options for identifying of landforms expanded. Nowadays, various remote sensing technologies are available to capture the surface terrain and elevation data.

- Stereophotogrammetry (elevation differences are derived by comparing a pair of optical images that target the same area and are acquired from slightly different positions)
- Radargrammetry (are based on stereoscopic pairs of amplitude SAR images acquired from the same side but with different incidence angles)
- SAR interferometry (InSAR; the elevation is derived from a pair of phase images acquired with the same polarization but taken from slightly different positions)
- Lidar (Sect. 10.3.3, *Lidar Sensors*)
- Radar altimeters (Sect. 10.3.3).

Based on these technologies, digital terrain models (DTMs), which represent the height of the terrain surface, and digital surface models (DSMs), which consider the height of the terrain surface as well as the heights of buildings, trees, etc., can be generated. DTMs and/or DSMs can be deployed to identify morphological features and events, for hydrological and sediment transport modeling, and for the estimation of soil erosion [212]. Elevation products generated by means of airborne lidar, InSAR, drone-based stereoscopic optical data, and terrestrial laser scanner data are characterized by high absolute height accuracy, so they can be utilized for the detection of terrain or surface changes.

Surfaces of outcropping rocks or bare soil can be identified via DTMs and DSMs, and their mineral compositions can then be investigated with sensors operating in the VIS to TIR range [213]. Due to their high spectral resolution and narrow bands, hyperspectral sensors well suited for discerning minerals with spectrally adjacent absorption features [214]. Hyperspectral data acquired in the VIS to SWIR range enable, for example, the detection of carbonate, clay, and iron [215], while TIR data are used to identify and to quantify carbonate, clay, quartz, feldspar, olivine, pyroxene, and micas possess diagnostic spectral features [216]. Due to the ground coverage of hyperspectral sensors, the spatial availability of hyperspectral data is limited. The resulting data gaps can be filled via the synergistic use of hyper- and multispectral data [217].

10.6.5 Forest Applications of Remote Sensing

In the context of the evaluation of aerial photographs, the prerequisites and objectives of RS in vegetation science are very different from those of RS in forestry (Table 10.8). Nevertheless, the evaluation methods and procedures used are similar in both cases. In the context of inventory, the characteristics used for the interpretation of aerial photographs are essentially the backscattering capacities of different wood species during the vegetation period, the phenology, the stage of plant development, and differences in crown shape and texture [211, 218].

The diversity of and disturbances in forest and land-use intensity are multidimensional and tremendously complex when considered spatiotemporally. RS approaches are suitable for ascertaining forest status and stress, forest disturbances, and resource limitations on forest diversity and health at different levels of biological organization (molecular, genetic, individual, species, populations, communities, biomes, ecosystems, and landscapes). This entails monitoring the trait, phylogenetic, taxonomic, structural, and functional diversity of forest vegetation with RS (Table 10.8).

1. Trait diversity: "Ecologists are increasingly looking at traits—rather than species—to measure the status, disturbances or health of ecosystems" [219]. Furthermore, different stress factors, processes, drivers, or resource limitations lead to changes in traits and to trait variations in forest ecosystems. Therefore, traits represent a crucial feature and indicator when attempting to understand and measure forestry. RS sensors on various platforms are able to monitor traits and trait variations [220] based on the principles of image spectroscopy across the electromagnetic spectrum from the visible to the microwave band [221].
2. Traits are defined as biochemical, physical, morphological, textural, structural, or functional characteristics of forest biotic components. Examples of biochemical-

biophysical traits include pigment content, chlorophyll *a* and *b*, α- and β-carotene, xanthophyll, and plant water content, while examples of physiological and functional traits include photosynthesis, chlorophyll fluorescence, and carbon sequestration [222]. "The ability to monitor forest health indicators with RS data depends on the following factors: (i) the characteristics of forest traits and the shape, density and distribution of forest traits in space and time, (ii) the spatial, spectral, radiometric, angular and temporal resolutions of RS sensors or multi-sensor systems, (iii) the choice of the modeling technique (classification or biophysical/chemical variable estimation) and entity representation (pixel-based or geographic object-based), as well as (iv) how well the RS algorithm and its assumptions fit the RS data and the plant traits and trait variations in forest ecosystems" [222].
3. The phylogenetic forest diversity is a measure of the length of the evolutionary pathways linked to a given set of forest plant taxa [223]. Phylogenetic and epigenetic processes lead to responses in the form of biochemical or morphological characteristics and adaptations in forest vegetation. The spectral fingerprint of forest plant species and the canopy is determined by their phylogenetic characteristics and epigenetic processes [224–227].
4. The taxonomic forest diversity describes the composition and configuration of various forest plant taxa. Taxonomically different forest plant species can be identified with RS if the spectral responses (which depend on the plant traits) of the individual forest species are different from one another. The identified of forest plant taxa can be further improved by analyzing time series [228].
5. The structural forest diversity and forest disturbances describe the composition and configuration of 2-D to 4-D forest structures. For example, a forest fire leads to changes in the structures and patterns of trees and their communities, resulting in a change in spectral response in the RS data [229]. Lidar sensors placed on airborne or UAV platforms can provide a range of forest inventory data and stand parameters, such as the basal area timber volume, the number of stems per hectare, the mean height, the mean height weighted by the basal area, the dominant height, the mean breast height, the mean diameter weighted by the basal area at breast height, the stand density index, the crown competition factor, and the relative density of trees [230].
6. The functional vegetation diversity refers to the diversity of realized ecological functions and processes in forest ecosystems. [231, 320] present an approach to monitoring functional diversity based on remotely sensed morphological and physiological forest traits. This approach is a fundamental step forward in mapping functional vegetation diversity using RS.

There are a huge number of recent and future satellite remote-sensing missions that are relevant to the assessment of forest inventory information and forest health using RS [323]. It must be said, however, that there is no single type of RS data or RS approach that is sufficient, comprehensive, flexible, and cost-effective enough to perform forest health monitoring of short-term to long-term processes at local to global scales. Therefore, it is imperative to develop and apply a multi-source forest diversity and health monitoring network (MUSO-FDH-MN) for linking multi-source data (close-range, airborne, and spaceborne RS data) as well as different in situ monitoring approaches [232].

10.6.6 Agricultural Applications of Remote Sensing

Anthropogenic management and land-use intensity can lead to changes in the range of species, plant traits, and plant community traits present, as well as to disturbances to the above- and belowground interactions of the soil with plant species [233]. Agricultural land covers a large fraction of the total land area of the Earth (51×10^6 km^2 of agricultural land out of a total land area of 149×10^6 km^2) [234]. Due to its extent and its importance in food production, agricultural land is a target for RS applications. Since agriculture is influenced by a variety of spatial and temporal environmental factors (e.g., diseases and the water content), rapid responses from farmers are required to ensure the stable production of high-quality food, feed, renewable resources, and bioenergy. Optical RS data, IR RS data, radar data, and lidar data have been successfully utilized to better understand the structures, processes, demands, dynamics, and land use intensity, as well as to the develop local and global agrarian management and/or biodiversity strategies [235]. Crop rotation can be captured by long-term time series of RS optical (Landsat, Sentinel-2) or radar [236] data.

The RS-based detection of vegetation, aboveground water networks, and land use is quite straightforward, in contrast to the detection of soil and geological structures, which is complicated by the presence of belowground components. These structures are estimated from a number of influencing factors that interact in a complex manner with each other, such as precipitation, temperature, grain-size distribution, water balance, humus content, and the edaphon [237]. Nevertheless, strategies to map soil properties and their variability whilst minimizing the terrestrial measuring effort (costs, personnel) through the use of aerial photographs were developed as far back as the 1950s and 1960s ([238], cited in [211]).

Processes and drivers such as the land-use intensity and management strategy influence the vegetation and thus plant traits such as the species composition and the sensitivity to soil solidification or to biochemical reactions against pesti-

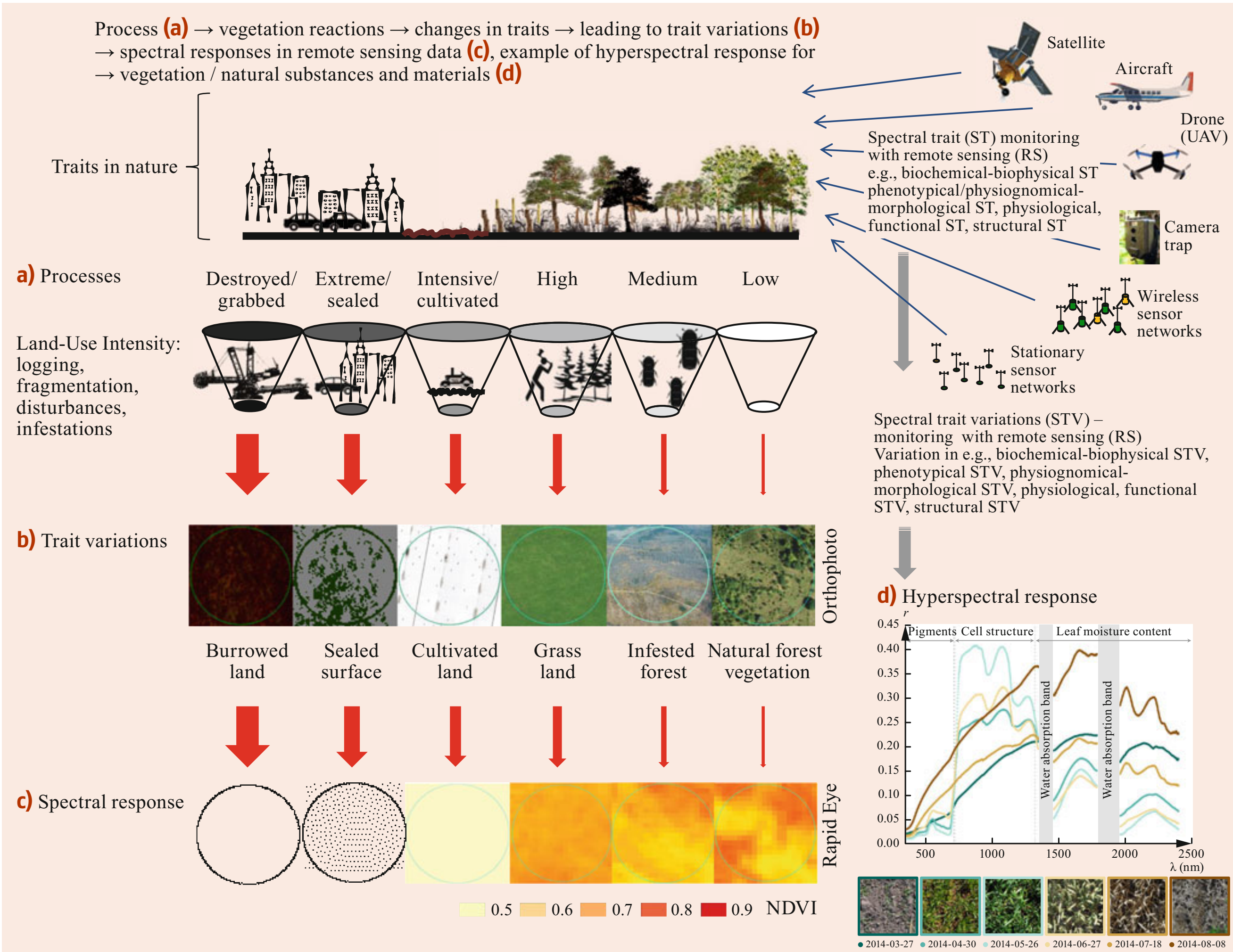


Fig. 10.8 Schema showing how the landscape characteristics influenced by (**a**) human impact depend on the land-use intensity and corresponding processes. Human activities cause (**b**) changes in traits and trait variations that can be (**c**) observed in the spectral response detected by close-range, airborne, and spaceborne RS systems. Trait variations that lead to changes in the hyperspectral response acquired over arable land are illustrated in (**d**) (adapted from [232])

cides. On the other hand, trait variations generate a change in the spectral response in RS data (see also Fig. 10.2). Furthermore, RS is successfully applied to deliver agrometeorological products [239], hydrological and soil parameters such as the mineralogy, C-org, clay texture, carbonate and iron contents, and soil moisture [37, 240–243], vegetation parameters [156], digital elevation or terrain models, area-wide land-use data, information on agricultural site status, and operative control over agronomic processes.

RS data are also deployed as an indicator of the intensity of grassland use [244]. Further, they enable the quantification of fertilizer application, the degree of mechanization, the frequency of harvesting and production [245], and the dynamics of arable land and pastures [246]. RS is a crucial tool for assessing the botanical composition, structure, quantity, and quality of temperate grasslands [247]. The integration of optical and thermal RS data into plant community and stand monitoring has also enabled significant advances in the detection of plant health and plant disease [248]. Additionally, numerous indicators such as fertilizer status, degree of intensification, biomass, and water management indicators have been developed in the field of precision farming [249, 250].

10.6.7 Remote Sensing of the Abiotic Diversity

Various abiotic factors influence the richness, abundance, and diversity of flora and fauna [251–253], energy and water vapor exchange at the biosphere–atmosphere interface [254], and characteristic landscape-scale patterns in soil microbial communities [255]. Thus, on the one hand, they codetermine interactions between abiotic and biotic environmental components at all spatiotemporal scales [256, 257]; on the other hand, they reflect changes in or disturbances of entities in the biotic and abiotic habitat as well as in ecosystem functions and services, as caused, for example, by rapid global changes and increased human intervention in ecosystems [233, 258–260]. The detection and identification of these changes and disturbances and a timely response to them require robust monitoring methods and statistical models that can monitor, describe, and predict biodiversity and its interactions with its abiotic compartments over space and time [259, 260].

Abiotic diversity in, for instance, terrestrial characteristics [213], soil [243], soil moisture, land surface temperature, water characteristics, and water quality [261] can be monitored by RS systems. The diversity of available sensors, such as field spectrometers, cameras, wireless sensor networks, tower installations, drones, airplanes, and satellites, permits monitoring at levels ranging from the local to the global level with high to low frequency at different spatiotemporal scales. Since abiotic diversity is caused by (among other factors) differences in the presence, distribution, frequency, and interactions of abiotic trait values, ecologists are increasingly focusing on trait investigations to quantify the status of ecosystems [219]. Abiotic traits are chemical, physical, morphological, textural, structural, or functional characteristics of abiotic components such as the land, soil, water, and atmosphere [262]. Some of them, the so-called "spectral abiotic properties" and their temporal and thus spectral changes, can be monitored by RSS approaches. To achieve this, spectroscopic methods that focus on the range from the VIS region to the MW region are deployed [220]. In the upcoming paragraphs, selected examples of RS-based monitoring approaches for abiotic trait diversity will be presented.

Terrestrial Diversity

Terrestrial diversity is influenced by diverse traits that are represented by DTMs and DSMs (Sect. 10.6.4). In combination with multitemporal and multi- or hyperspectral RS data, DTMs and DSMs are used to extract biodiversity features such as the vertical vegetation structure, biomass, primary production, microclimate, and precipitation of species [263] as well as to quantify, model, and monitor plant and animal species distributions [233]. DTMs and DSMs with high spatial resolution are deployed to model movement and migration behavior and to describe the microclimates of animal and plant species [264].

Soil Diversity

The diversity of soils is another abiotic factor. It determines in particular the biodiversity and thereby the ecology of an area by influencing the local climate conditions and the vegetation type and structure, for example. Both topographic and soil formations are influenced by the characteristics of the bedrock (e.g., mineralogical composition, erodibility; [233, 265]). RS-based mapping of surface mineralogy and lithology by hyper- and multispectral sensors operating in the VIS to TIR range can significantly contribute to an area-wide understanding of land-cover patterns and changes in them (Sect. 10.6.4, [213, 242, 266]).

Abiotic soil-related factors such as nutrient availability affect the biodiversity, and the biodiversity reflects the spatial heterogeneity of the soil [213, 267]. Therefore, RS images of bare soil and vegetation depict the prevailing soil properties directly and indirectly, respectively [213]. The reflectance of bare soils can be directly captured by passive or optical systems and enables the prediction of parameters such as the soil texture, soil organic carbon, iron content, soil salinity, carbonates, soil moisture, and even soil type classification. In contrast, active systems operating in the MW range (e.g., SAR or lidar) are mainly used to detect surface properties or to estimate the soil moisture content [268].

Soil Moisture Diversity

Soil moisture constitutes 0.001% of all the freshwater stored globally [269]. Due to its temporal and spatial dynamics and its involvement in evaporation, transpiration, infiltration, and groundwater replenishment processes, soil moisture affects the composition, structure, functionality, and dynamics of biodiversity [269–271]. Moreover, soil moisture controls biosphere–pedosphere and process–pattern interactions for diverse soil, hydrological, biological, climate, and other ecosystem processes [271] and accounts for 65% of the variation in ecosystem multifunctionality [270]. Therefore, comprehensive knowledge about the spatial and temporal heterogeneity of soil moisture is required for the modeling of water and energy fluxes, nutrient transport, and matter turnover within soil landscape systems [272, 273]. This knowledge about soil moisture patterns can be derived directly for areas with little or no vegetation and indirectly for vegetated areas from measurements acquired with active and passive MW sensors as well as optical sensors, using vegetation traits as a proxy.

10

MW sensors detect increases and decreases in the proportion of water in the soil via increases and decreases in the electrical conductivity of the soil, respectively [19]. This approach is based on the assumption that soil is composed of water ($\varepsilon_{\text{water}} \approx 81$), air ($\varepsilon_{\text{air}} = 1$), and soil particle ($\varepsilon_{\text{soil particle}} \approx 2$–$5$) phases and the proportion of each phase can change [274]. In addition, the detected signal can be affected by sensor characteristics (e.g., polarization, incidence angle, wavelength, and spatial as well as temporal resolution) and surface characteristics (e.g., soil surface roughness and vegetation cover with different geometries, densities, and water contents; [19, 275]).

Active microwave sensors such as radars and scatterometers record backscattered MW radiation in the form of normalized backscattering coefficients or normalized radar cross-sections [321]. These direct soil moisture detection methods are independent of the illumination and weather conditions [276] and work with a higher radiation budget than passive systems, leading to shorter integration times and therefore spatial resolutions in the range of a meter. The number of different scatterers that contribute to the backscattered radiance increases with the appearance of vegetation and hampers the extraction of the soil moisture signal, especially for shorter wavelengths. This is why acquisitions with longer wavelengths, such as those in the L band, are preferred for soil moisture retrieval in areas covered by vegetation [277]. Decomposition methods are deployed to extract the part of the backscatter signal caused by the moisture of the soil [278]. Empirical (including machine learning, [279–281], semiempirical [282–284], and physically-based [19, 285, 286]) algorithms are utilized to estimate the soil moisture. Another method for retrieving the soil moisture dynamics from active MW sensor measurements is differential SAR interferometry [287–289].

Passive microwave sensors such as radiometers record the MW radiation emitted from the Earth in the form of the brightness temperature T_{b}, which is the product of the emissivity ε and the physical temperature T_{object} of the targeted object. The retrieval of the soil moisture from the detected brightness temperature T_{b} requires knowledge of the physical temperature T_{object}. The emissivity ε of a smooth surface, which is a function of the incidence angle and the electrical conductivity of the soil, can be predicted using the Fresnel reflection equations. Theoretically, increasing the soil roughness leads to a decrease in the reflectivity and an increase in the thermal emissivity of the soil. Practical experience shows that the behavior of rough soil is probably more affected by the distribution of water in the top soil than the geometry of the soil surface roughness. The low radiation budget available for this method necessitates long passive sensor integration times, resulting in spatial resolutions of kilometers. The occurrence of vegetation affects the land surface emission signal such that soil moisture estimation is only possible up to a vegetation water content of $5\,\mathrm{kg/m^2}$, even if radiation in the L band is collected [290, 291].

Various optical and thermal sensors allow for direct and indirect measurements of soil moisture [292–294]. The suitability of a sensor for the detection of soil moisture depends mainly on its spatial, spectral, and temporal resolution. The direct retrieval of soil moisture from optical and thermal RS data requires bare soils and a massive amount of in situ soil moisture data to account for the high spatial and temporal variability of the parameter. Soil moisture retrieval from optical and thermal RS data for vegetation-covered soils is only possible indirectly. In this case, biochemical, morphological, physiological, or functional characteristics of vegetation or spectral plant traits are used as a proxy for the soil moisture determination [295].

Land Surface Temperature

The land surface temperature (LST) is influenced by the radiative, thermal, and hydraulic properties of the soil–plant–atmosphere system [296] and provides information about the water and energy exchange in this system from the local to the global scale [297, 298]. Therefore, LST can be utilized to describe the photosynthetic activity of the vegetation [299]) and estimate the evapotranspiration [300]. Remotely sensed TIR radiances allow for the derivation of the LST if the acquired radiance is radiometrically calibrated and corrected for the influences of the surface emissivity, clouds, and atmospheric conditions [301].

10.6.8 Urban Remote Sensing

The RS of urban areas is important but challenging due to the small-scale heterogeneous pattern of buildings, traffic infrastructure, gardens, and other types of land cover. The following examples may illustrate the multiplicity of detail involved: different roof coverings, vegetation, water, small-scale temperature and climatic changes, rapid changes in light and shade, and many building types, such as family houses with gardens, apartment buildings, and urban structures (e.g., big buildings in older city centers or scattered buildings in urban-rural transition areas). This results in several specific requirements for the spatial, temporal, and spectral resolutions of RS sensors.

Applications of RS are manifold and concern the monitoring of urban structures at global and regional scales (e.g., the Global Urban Footprint; [302, 303]). Works at a regional scale include the documentation of a city's rural surroundings, the mapping of sealed areas to aid the sewage system planning, temperature and climate analyses, and census assessment.

10.6.9 Archeological Applications of Remote Sensing

RS provides large-scale observations in short repeated cycles. As a result, good overview information with almost simultaneous observation of sites with different sensor types is aimed for. This requires favorable monitoring that is optimally adapted to the task, e.g. by selecting a suitable sensor combination or optimal recording times. The demands for archeology may differ from those for other applications, since conditions that are usually disadvantageous—such as a low Sun elevation—may be desirable, as it can amplify the detectability of minor landscape and soil undulations.

In addition to the characteristic fingerprints of archeological targets, RS methods can provide information on altitudes, distances to water or between sites or cities, and corridors and traffic routes. Such information can be used to predict the locations of potential archeological sites [304].

Aerial photography was the first form of remote support for archeology. It was used for location recognition and archeological mapping, as it made it possible to identify patterns and details of features that were not easily visible in ground surveys [211, 305, 306]). With the increased availability and provision of cost-effective satellite data as well as improvements in the technological parameters of these data (e.g., the spatial, spectral, and temporal resolution), satellite data were increasingly introduced into archeology. For example, these data were successfully used to prospect for and to recognize and classify settlement structures [307, 308]. In particular, lidar RS has led to new findings. Lidar measurements have supported the detection of pre-Columbian settlement patterns and land uses [309], the localization of building structures [310] and the creation of location maps [311]. Additionally, UAVs and minimized sensors offer new possibilities for flexible application in terms of detection optimization.

However, it has been shown that the use of RS methods (e.g., ground-based RS, photogrammetry, satellite RS, and lidar) along with other technologies (e.g., GIS) or open-source data and civic approaches can provide valuable information, especially when they are used in a coordinated and complementary way [312, 313].

10.6.10 Application of Remote Sensing to Hazard and Disaster Information Gathering

A hazard is a natural (e.g., a storm, heatwave, freeze, landslide, drought, flood, epidemic, or infestation), or human-induced/made (e.g., a fire, explosion, or collision) situation or phenomenon that may have a negative impact on human life, infrastructure, or environmental resources. In contrast, a disaster is considered by the International Federation of Red Cross and Red Crescent Societies [314] to be the sum of vulnerability and hazard divided by the capacity to provide assistance. Vulnerability is the predisposition to suffer damage due to an external event. Capacity describes the personal and material resources available to cope with an impact on the social life and functioning of a society.

A disaster can be defined as "a serious disruption of the functioning of a community or a society involving widespread human, material, economic or environmental losses and impacts, which exceeds the ability of the affected community or society" (at a local, regional, or global scale) "to cope using its own resources" [315]. Reference [316] provide a classification of hazards and disasters according to their drivers and causes.

RS data have been successfully used to monitor hazards or disasters. Examples include the detection of the reactor disaster in Chernobyl on May 1, 1986 by SPOT (Satellite Pour l'Observation de la Terre), the monitoring of the catastrophic forest fires in California by the Nasa Terra and Aqua satellites in 2009 and in 2018, the provision of information on a series of tsunamis along coasts in the Indian Ocean caused by a submarine earthquake in 2004, and the reactor disaster following a tsunami in Fukushima in 2011.

Against this background, various signatories such as NOAA, USGS, ISRO, JAXA, and DLR have initiated an international charter for space and major disasters in order to provide immediate operational data and information that can assist decision-makers and auxiliary forces in conflict areas in identifying and mapping most affected areas or the areas in which auxiliary facilities should be installed, in optimizing limited available local or regional resources or supplies from state and aid organizations, and in managing limited resources (auxiliary staff, etc.).

Given optimal coordination of the data provision as well as operational processing and fusion with additional data and geoinformation, RS is a helpful tool for the real-time or near-real-time provision of current spatial information.

10.7 Recent and Future Developments

There are a wide range of possible fields of application for RS. The most important market drivers in the last few years have been (i) the availability of high-resolution RS data, (ii) easier access to RS data and products for users, (iii) lower data costs, (iv) the development of the GIS sector, and (v) the development of new processing and evaluation methods.

With the implementation of Copernicus (formerly GMES), a joint initiative of the European Union and the European Space Agency, concrete steps were taken to develop the European geoinformation market by consolidating the space and ground segments as well as the use of RS data. Other factors that have driven the development of an international geoinformation market include (i) the increasing number of nations operating RS missions, (ii) the entry of commercial market participants into, e.g., the space, ground, and distribution segments, (iii) the increasing number of commercial information products or services that are based on them, (iv) the rapid development of the telecommunications and advertising market, and (v) the rapid development of computer technology, software technology, and Internet technology.

These market drivers have been accompanied by free access to RS data delivered by USGS (Landsat), DLR/RapidEye (RapidEye), DLR/Astrium (TerraSAR), and ESA (Sentinel-1, -2, and -3) for, e.g., the scientific development of automated processors and processing chains.

All of the activities and efforts that are being directed into the further development of RS are aimed at making improvements to RS. This objective is based on the desire to maintain data continuity while at the same time ensuring data integrity, data correctness, data trustworthiness, and data quality. This means that the provision of data by the data suppliers and the data requirements of the users must be brought into line.

An essential aspect of technical developments in sensor science and computer science is the increasing miniaturization of electronic components, which has also led to a reduction in the use of electrical energy in such components. This miniaturization can be expected to continue over the next few years. This will influence the development of very small and small satellites and the development of optical and microwave sensors, so long as there are no technical or construction limits on this development. Miniaturization will accommodate both Earth exploration and RS through the development of miniature aircraft and UAVs.

A major field of development in RS is that of RS UAVs, which enable ad-hoc deployment that approaches the user's need for information. This development will lead to largely autonomous controlled systems that can automatically process a planned flight guidance, with the systems also increasingly making corrections independently. In addition, the systems are increasingly being optimized so that a comparatively large payload can be carried during a long flight. An essential aspect of this development will be the development of novel batteries and the miniaturization of sensors and auxiliary systems. Through the development of end-to-end chains, data will be available to the user immediately after data recording. This is an important field of work in the development of fast processing algorithms.

Since UAVs can very quickly record large amounts of data due to the large overlap of recorded images required, great technological demands are placed on the further processing of the data. Accordingly, processing structures similar to those already developed in satellite-based RS may become necessary. These include effective flight-planning systems, carrier–sensor systems, regional reception systems for the data volumes generated, data processing chains, and control systems for the processors.

Techniques and technologies for minimizing weather-related limitations of optical RS were developed at a very early stage. For example, several satellites equipped with radar sensors were developed, which are largely independent of the weather conditions and independent to the illumination of the Earth surface by sun, or were equipped with swivelling sensors onboard (SPOT, [39]) or several satellites with optical sensors placed on an orbit, making it possible to observe a point on the Earth's surface at shorter intervals (RapidEye [39]).

Recent developments in the field of satellite RS show that data are being recorded at ever higher spatial and spectral resolutions. The ever-larger amounts of data generated have to be transmitted from the satellite to ground stations. Whereas these data were previously transmitted via the X band, higher data rate transmission must now occur via the K_a band. Thus, the advantage of largely weather-dependent data transmission via the X band is eliminated in favor of higher data transmission density via the K_a band. In order to be able to realize a uniform data stream in this case, geostationary relay satellites that will receive the data from orbiting satellites and transmit them to the ground stations on Earth are proposed. In this case, the design of the ground-based receiving stations is less complicated, as they do not have to follow the movement of the satellite, whereas they did before [317].

Another development that serves to minimize the amount of data to be transmitted is onboard processing. The received raw data are evaluated and processed directly onboard the satellite according to previously defined algorithms. The evaluation results are sent to a ground station, from where the data are distributed.

References

1. Mulders, M.A.: Remote sensing in soil science. In: Mulders, M.A. (ed.) Developments in Soil Science, vol. 15, p. 370. Elsevier, Amsterdam (1987)
2. Jensen, J.R.: Remote Sensing of the Environment. An Earth Resource Perspective. Prentice Hall, Upper Saddle River, p. 592 (2000)
3. Albertz, J.: Einführung in die Fernerkundung. Grundlagen der Interpretation von Luft- und Satellitenbildern. Wissenschaftliche Buchges., Darmstadt, p. 254 (2009)
4. Liang, S., Strahler, A.: Retrieval of surface BRDF from multiangle remotely sensed data. Remote Sens. Environ. **50**, 18–30 (1994)
5. Ulaby, F.T., Moore, R.K., Fung, A.K.: Microwave Remote Sensing, Vol. I: Microwave Remote Sensing Fundamentals and Radiometry. Addison-Wesley, London, p. 456 (1981)
6. Gupta, R.P.: Remote Sensing Geology. Springer, Berlin (2018)
7. Wooster, M.J., Roberts, G., Smith, A.M.S., Johnston, J., Freeborn, P., Amici, S., Hudak, A.T.: Thermal remote sensing of active vegetation fires and biomass burning events. In: Thermal Infrared Remote Sensing, pp. 347–390. Springer, Dordrecht (2013)
8. EUMeTrain: Forest Fires (2012). http://www.eumetrain.org/data/3/30/print_5.htm#page_1.1.0, Accessed 10 Dec 2018
9. Rohde, R.A.: Radiation Transmitted by the Atmosphere (Wikipedia 2007) (2007). https://en.wikipedia.org/wiki/File:Atmospheric_Transmission.png, Accessed 23 Nov 2018
10. Boyd, D.S.: Physical basis of remote sensing. In: Atkinson, P.M. (ed.) Geoinformatics, Vol. I Encyclopedia of Life Support. pp. 184–208. EOLSS Publishers, Oxford (2008)
11. Kraus, K., Schneider, W.: Fernerkundung – Physikalische Grundlagen und Aufnahmetechniken vol. 1. Dümmler, Bonn, p. 300 (1988)
12. Long, M.W.: Radar Reflectivity of Land and Sea. Lexington Books, London, p. 306 (1975)
13. Barrett, E.C., Curtis, L.F.: Introduction to Environmental Remote Sensing. Chapman Hall, London, p. 352 (1982)
14. Combs, A.C., Weickmann, H.K., Mader, C., Tebo, A.: Application of infrared radiometers to meteorology. J. Appl. Meteorol. **4**, 253–262 (1965)
15. Sahu, K.C.: Textbook of Remote Sensing and Geographical Information Systems. Atlantic, New Delhi, p. 503 (2008)

16. van de Hulst, H.C.: Light Scattering by Small Particles. Wiley, New York, p. 470 (1957)
17. Mishchenko, M.I., Travis, L.D., Lacis, A.A.: Scattering, Absorption, and Emission of Light by Small Particles. Cambridge Univ. Press, Cambridge, p. 445 (2002)
18. Mo, T., Schmugge, T.J., Jachson, T.J.: Calculations of radar backscattering coefficient of vegetation-covered soils. Remote Sens. Environ. **15**, 119–133 (1984)
19. Ulaby, F.T., Moore, R.K., Fung, A.K.: Microwave Remote Sensing, Vol. II: Active and Passive, Radar Remote Sensing and Surface Scattering and Emission Theory. Addison-Wesley, London, p. 609 (1982)
20. von Hippel, A.R.: Dielectric Materials and Applications. MIT Technology Press, New York, p. 438 (1954)
21. Nelson, S., D. Lindroth, Blake, R.: Dielectric properties of selected minerals at 1 to 22 GHz, Geophysics, 54(4), 1334–1349 (1989)
22. Ulriksen, C.P.F.: Application of Impulse Radar to Civil Engineering. Dissertation. Univ. Lund, Lund, p. 175 (1982)
23. Myllys, M., Simojoki, A.: Calibration of time domain reflectometry (TDR) for soil moisture measurements in cultivated peat soils. Suo **47**(1), 1–6 (1996)
24. de Vries, D.A.: Thermal properties of soils. In: van Wijk, W.R. (ed.) Physics of Plant Environment, pp. 210–235. North Holland, Amsterdam (1963)
25. Arenson, L.U., Colgan, W., Marshall, H.-P.: Physical, thermal, and mechanical properties of snow, ice, and permafrost. In: Snow and Ice-Related Hazards, Risks, and Disasters, pp. 35–75. Academic Press, New York (2015) https://doi.org/10.1016/B978-0-12-394849-6.00002-0. Chap. 2
26. Mackowski, D.W.: MSTM Version 3.0 (2013). http://www.eng.auburn.edu/users/dmckwski/scatcodes/
27. Chandrasekhar, S.: Radiative Transfer. Dover, New York, p. 416 (1960)
28. Popp, T.: Korrektur der atmosphärischen Maskierung zur Bestimmung der spektralen Albedo von Landoberflächen aus Satellitenmessungen. Dissertation. Univ. München, München, p. 137 (1993)
29. Vermote, E.F., Vermeulen, A.: Atmospheric Correction Algorithm: Spectral Reflectances (MOD09). MODIS Algorithm Technical Background Document. Version 4.0, April 1999 (1999). https://eospso.nasa.gov/sites/default/files/atbd/atbd_mod08.pdf, Accessed 11 Feb 2019. NASA contract NAS5-96062
30. Erbertseder, T.: Quantifizierung von atmosphärischen Einflüssen auf NOAA-AVHRR NDVI-Daten und deren Korrektur. Diploma thesis. Ludwig-Maximilians-Univ. München, München (1998)
31. Holzer-Popp, T., Bittner, M., Borg, E., Dech, S., Erbertseder, T., Fichtelmann, B., Schroedter, M.: Das Automatische Atmosphärenkorrekturverfahren "DurchBlick". In: Blaschke, T. (ed.) Fernerkundung und GIS: Neue Sensoren – Innovative Methoden, pp. 78–87. Wichmann, Heidelberg (2001)
32. Holzer-Popp, T., Bittner, M., Borg, E., Dech, S., Erbertseder, T., Fichtelmann, B., Schroedter, M. (2004): Process for Correcting Atmospheric Influences in Multispectral Optical Remote Sensing, Eur. Patent 1091188 B1
33. Truckenbrodt, S.C., Baade, J.: Gebesee Database for the Enhancement of Crop Monitoring Applications: Evolution of Plant Physiology, Soil Moisture, Surface Reflectance and Atmospheric Conditions on the Agricultural Gebesee Test Site (Central Germany) in 2013 and 2014 (2017) https://doi.org/10.1594/PANGAEA.874251
34. Truckenbrodt, S.C., Schmullius, C.C.: Seasonal evolution of soil and plant parameters on the agricultural Gebesee test site: A database for the set-up and validation of EO-LDAS and satellite-aided retrieval models. Earth Syst. Sci. Data **10**(1), 525–548 (2018)
35. Robinson, I., MacArthur, A.: The Field Spectroscopy Facility Post Processing Toolbox User Guide, Post Processing Spectral Data in MATLAB. Univ. Edinburgh, Edinburgh, p. 24 (2011)
36. Vermote, E.F., Tanré, D., Deuzé, J.L., Herman, M., Morcrette, J.J.: Second simulation of the satellite signal in the solar spectrum, 6S: An overview. IEEE Trans. Geosci. Remote Sens. **35**(3), 675–686 (1997)
37. Blasch, G., Spengler, D., Itzerott, S., Wessolek, G.: Organic matter modeling at the landscape scale based on multitemporal soil pattern analysis using RapidEye data. Remote Sens. **7**(9), 11125–11150 (2015)
38. Ehlers, M.: Neue Sensoren und Methoden in der Fernerkundung: Zwischen hochauflösend und hyperspektral. In: Pixel und Profile – Fernerkundung in den Bodenwissenschaften, pp. 7–16. Fachhochschule Osnabrück, Osnabrück (2005). https://www.mabga.de/fileadmin/HSOS/Studium/Studienangebot/Studiengaenge/Masterstudiengaenge/AuL/Boden__Gewaesser__Altlasten/Dokumente/2005_Tagungsband_Pixel_Profile.pdf (last access: 2 October 2019)
39. Kramer, H.-J.: Observation of the Earth and Its Environment. Survey of Missions and Sensors, 4th edn. Springer, Berlin, p. 1510 (2002)
40. Campbell, J.B.: Introduction to Remote Sensing. Taylor Francis, London, p. 654 (2002)
41. Westin, T.: Geometric modelling of imagery from the MSU-SK conical scanner. Bull. Soc. Franç. Photogramm. Télédétect. **159**(3), 55–58 (2000)
42. Ulaby, F.T., Long, D.G.: Microwave Radar and Radiometric Remote Sensing. Univ. Michigan Press, Ann Arbor, p. 984 (2014)
43. Cracknell, A.P., Hayes, L.: Introduction to Remote Sensing. CRC, Boca Raton, p. 335 (2007)
44. de Loor, G.P.: Introduction and some general aspects of image formation in radar. Remote Sens. Rev. **1**(1), 3–18 (1983)
45. Wise, P.: Spaceborne Remote Imagery – Its acquisition, processing and cartographic applications. Cartography **18**(1), 9–20 (1989)
46. Kalkuhl, M.: Modulare Simulation komplexer SAR-Szenarien: Signalgenerierung, Positionsschätzung und Missionsplanung. Dissertation. Univ. Siegen, Siegen, p. 283 (2009)
47. Hein, A.: Verarbeitung von SAR-Daten unter besonderer Berücksichtigung interferometrischer Anwendungen. Dissertation. Univ. Siegen, Siegen (1998)
48. Goodman, J.W.: Some fundamental properties of speckle. J. Opt. Soc. Am. **66**(11), 1145–1150 (1976)
49. Lee, J.S., Jurkevich, I., Dewaele, P., Wambacq, P., Oosterlinck, A.: Speckle filtering of synthetic aperture radar images: A review. Remote Sens. Rev. **8**, 313–340 (1994)
50. van Zyl, J.J.: Radar Remote Sensing for Earth and Planetary Sciences. Jet Propulsion Laboratory, Pasadena (2007). http://www.its.caltech.edu/~ee157/lecture_note/Radar.pdf (last access: 10 February 2019)
51. COAPS (Center for Ocean-Atmospheric Prediction Studies): Scatterometry & Ocean Vector Winds. Satellite Studies (2018). https://coaps.fsu.edu/scatterometry/about/overview.php, Accessed 28 Nov 2018
52. Winker, D.M., Vaughan, M.A., Omar, A., Hu, Y., Powell, K.A., Liu, Z., Hunt, W.H., Young, S.A.: Overview of the CALIPSO mission and CALIOP data processing algorithms. J. Atmos. Ocean. Technol. **26**(11), 2310–2323 (2009)
53. Winker, D.M., Pelon, J., Coakley Jr., J.A., Ackerman, S.A., Charlson, R.J., Colarco, P.R., Flamant, P., Fu, Q., Hoff, R.M., Kittaka, C., Kubar, T.L., Le Treut, H., Mccormick, M.P., Mégie, G., Poole, L., Powell, K., Trepte, C., Vaughan, M.A., Wielicki, B.A.: The CALIPSO mission: A global 3D view of aerosols and clouds. Bull. Am. Meteorol. Soc. **91**(9), 1211–1229 (2010)

54. Illingworth, A.J., Barker, H.W., Beljaars, A., Ceccaldi, M., Chepfer, H., Clerbaux, N., Cole, J., Delanoë, J., Domenech, C., Donovan, D.P., Fukuda, S., Hirakata, M., Hogan, R.J., Huenerbein, A., Kollias, P., Kubota, T., Nakajima, T., Nakajima, T.Y., Nishizawa, T., Ohno, Y., Okamoto, H., Oki, R., Sato, K., Satoh, M., Shephard, M.W., Velázquez-Blázquez, A., Wandinger, U., Wehr, T., van Zadelhoff, G.-J.: The EarthCARE satellite: The next step forward in global measurements of clouds, aerosols, precipitation, and radiation. Bull. Am. Meteorol. Soc. **96**(8), 1311–1332 (2015)
55. Stoffelen, A., Pailleux, J., Källén, E., Vaughan, J.M., Isaksen, L., Flamant, P., Wergen, W., Andersson, E., Schyberg, H., Culoma, A., Meynart, R., Endemann, M., Ingmann, P.: The atmospheric dynamics mission for global wind field measurement. Bull. Am. Meteorol. Soc. **86**(1), 73–87 (2005)
56. Ehret, G., Flamant, P., Amediek, A., Ciais, P., Gibert, F., Fix, A., Kiemle, C., Quatrevalet, M., Wirth, M.: The French-German climate monitoring initiative on global observations of atmospheric methane. In: 25th Int. Laser Radar Conf. pp. 1348–1351. (2010)
57. Konecny, G.: Small satellites – A tool for Earth observation? – The classification of satellites according to mass is usually as follows. ISPRS Archives **35**(B4), 580–582 (2004)
58. Allain, R.: What's so Special About Low Earth Orbit? (2015). https://www.wired.com/2015/09/whats-special-low-earth-orbit/, Accessed 13 Nov 2018
59. Hillhouse, J.D.: Sun Synchronous Orbits for the Earth Solar Power Satellite System (1999). http://design.ae.utexas.edu/mission_planning/mission_resources/orbital_mechanics/Sun_Synchronous_Orbits.pdf, Accessed 30 Jan 2019
60. Hoffmann, E.J.: Space communications. In: Pisacane, V.L. (ed.) Fundamentals of Space Systems, pp. 565–599. Oxford Univ. Press, Oxford (2005)
61. Holdrige, M.E.: Space mission operations. In: Pisacane, V.L. (ed.) Fundamentals of Space Systems, pp. 754–771. Oxford University Press, Oxford (2005)
62. Montenbruck, O., Gill, E.: Satellite Orbits: Models, Methods and Applications. Springer, Berlin, p. 369 (2012)
63. NASA: NASA Earth Observatory. Three Classes of Orbit (2009). https://earthobservatory.nasa.gov/Features/OrbitsCatalog/page2.php, Accessed 13 Nov 2018
64. Qiao, L., Rizos, C., Dempster, A.G.: Satellites orbit design for the Australian Garada Project. In: Proc. 24th Int. Tech. Meet. (ION GNSS 2011), pp. 1075–1081. (2011)
65. Reiniger, K., Diedrich, E., Mikusch, E.: Aspects of ground segment design for Earth observation missions (1 August 2006). In: Summer School Alpbach 2016 "Monitoring of Natural Hazards from Space", p. 53. (2006)
66. Baret, F., Buis, S.: Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems. In: Liang, S. (ed.) Advances in Land Remote Sensing, pp. 173–201. Springer, Dordrecht (2008)
67. Clark, P.E., Rilee, M.L.: Remote Sensing Tools for Exploration – Observing and Interpreting the Electromagnetic Spectrum. Springer, New York, p. 346 (2010)
68. USGS: Landsat 8 (L8) Data Users Handbook, LSDS-1574, Version 3.0, October 2018 (2018). https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/LSDS-1574_L8_Data_Users_Handbook.pdf, Accessed 31 Jan 2019
69. Fichtelmann, B., Borg, E.: A new approach of geo-rectification for time series satellite data based on a graph-network. Lect. Notes Comput. Sci. **9158**, 217–232 (2015)
70. Fichtelmann, B., Borg, E., Günther, A., Guder, H.-G., Reimer, R., Wolff, E., Damerow, H.: Chlorophyll map of the Baltic Sea – The results of an automatic processing chain from data reception up to archiving and delivery of value-added products. In: Bruzzone, L., Smith, P. (eds.) European Commission – Joint Research Centre – Special Publications I.01.122, p. 50. (2001)
71. Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G.: High-resolution global maps of 21st-century forest cover change. Science **342**(6160), 850–853 (2013)
72. Fichtelmann, B., Günther, K.P., Borg, E.: Adaption of a Self-Learning Algorithm for Dynamic Classification of Water Bodies to SPOT VEGETATION Data. Lect. Notes Comput. Sci. **9158**, 177–192 (2015)
73. Liang, S. (ed.): Quantitative Remote Sensing of Land Surfaces. Wiley, Hoboken (2004)
74. Schwarz, J., Maass, H., Bettac, H.-D.: Data reception of small satellites in a multimission scenario. Acta Astron. **46**(2/6), 221–228 (2000)
75. Schwarz, J.: Highrate data acquisition system for Earth observation satellites. In: Proc. Eur. Telem. Conf. Exhib. Data Rec. Storage, pp. 122–131. (2002)
76. Böhme, C., Bouwer, P.J., Prinsloo, M.: Real-time stream processing for active fire monitoring on Landsat 8 direct reception data. ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **40**(7), 765–770 (2015)
77. Krause, D., Schwarz, E., Voinov, S., Damerow, H., Tomecki, D.: Sentinel-1 near real-time application for maritime situational awareness. CEAS Space J. **11**(1), 45–53 (2019)
78. Haralick, R.M.: Automatic remote sensor image processing. In: Rosenfield, A. (ed.) Digital Picture Analysis, pp. 5–63. Springer, New York (1976)
79. Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.-K.: A Landsat surface reflectance dataset for North America, 1990–2000. IEEE Geosci. Remote Sens. Lett. **3**(1), 68–72 (2006)
80. Irish, R.R.: LANDSAT 7 automatic cloud cover assessment. Proc. SPIE **4049**, 348–355 (2000)
81. Biasutti, R.: Cloud cover evaluation. In: 8th LANDSAT Tech. Work. Group Meet. p. 10. (2000)
82. Borg, E., Fichtelmann, B., Asche, H.: Assessment for remote sensing data: Accuracy of interactive data quality interpretation. Lect. Notes Comput. Sci. **6783**, 366–375 (2011)
83. Scaramuzza, P., Micijevic, E., Chander, G.: SLC Gap-Filled Products – Phase One Methodology. SLC Gap-Fill Methodology, 18 March 2004 (2004). https://landsat.usgs.gov/filling-gaps-use-scientific-analysis#2, Accessed 22 Nov 2018
84. USGS: Landsat Missions. Landsat 7 Data Users Handbook – Appendix A, Known Issues (2018). https://landsat.usgs.gov/landsat-7-data-users-handbook-appendix-a, Accessed 22 Nov 2018
85. Richards, J. A., Jia, X.: Remote sensing digital image analysis: An introduction, 4th edn. Springer-Verlag, Berlin, Germany, p. 439 (2006)
86. Kaur, S.: Noise types and various removal techniques. Int. J. Adv. Res. Electron. Commun. Eng. **4**(2), 226–230 (2015)
87. Al-amri, S.S., Kalyankar, N.V., Khamitkar, S.D.: A comparative study of removal noise from remote sensing image. Int. J. Comput. Sci. Issues **7**(1), 32–36 (2010)
88. Bhosale, N.P., Manza, R.R.: A review on noise removal techniques from remote sensing images. In: Nat. Conf. Radhai College (2013) https://doi.org/10.13140/RG.2.1.1104.6168. CMS-13
89. Townshend, J.R.G., Huang, C., Kalluri, S.N.V., Defries, R.S., Liang, S., Yang, K.: The impact of misregistration on change detection. IEEE Trans. Geosci. Remote Sens. **30**(5), 1054–1060 (1992)

90. Dai, X., Khorram, S.: The effects of image misregistration on the accuracy of remotely sensed change detection. IEEE Trans. Geosci. Remote Sens. **36**(5), 1566–1577 (1998)
91. Verbyla, D.L., Boles, S.H.: Bias in land cover change estimates due to misregistration. Int. J. Remote Sens. **21**, 3553–3560 (2000)
92. Hill, L.L.: Georeferencing: The Geographic Associations of Information. MIT Press, London, p. 280 (2006)
93. Hackeloeer, A., Klasing, K., Krisp, J.M., Meng, L.: Georeferencing: A review of methods and applications. Annals GIS **20**(1), 61–69 (2014)
94. Richards, R.A.: Remote Sensing Digital Image Analysis – An Introduction. Springer, Berlin, Heidelberg, p. 494 (2013)
95. Toutin, T.: Geometric processing of remote sensing images: Models, algorithms and methods. Int. J. Remote Sens. **25**(10), 1893–1924 (2004). Review article
96. Ries, C.: Ein allgemeiner Ansatz zur Georeferenzierung von multispektralen Flugzeugscanneraufnahmen. Dissertation. TU Wien, Wien (2004)
97. Brown, L.G.: A survey of image registration techniques. ACM Comput. Surv. **24**(4), 325–376 (1992)
98. Bannari, A., Morin, D., Bénié, G.B., Bonn, F.J.: A theoretical review of different mathematical models of geometric corrections applied to remote sensing images. Remote Sens. Rev. **13**, 27–47 (1995)
99. Fonseca, L.M.G., Manjunath, B.S.: Registration techniques for multisensor remotely sensed imagery. Photogramm. Eng. Remote Sens. **62**(9), 1049–1056 (1996)
100. Bartoli, G.: Image Registration Techniques: A Comprehensive Survey (2007). http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.471.1134&rank=1, Accessed 25 Jan 2019
101. Hoja, D., Schneider, M., Müller, R., Lehner, M., Reinartz, P.: Comparison of orthorectification methods suitable for rapid mapping using direct georeferencing and RPC for optical satellite data. ISPRS Archives **37**(B4), 1617–1623 (2008)
102. Norjamäki, I., Tokola, T.: Comparison of atmospheric correction methods in mapping timber volume with multitemporal Landsat images in Kainuu, Finland. Photogramm. Eng. Remote Sens. **73**(2), 155–163 (2007)
103. Richter, R., Schläpfer, D.: Atmospheric/Topographic Correction for Satellite Imagery: ATCOR-2/3 User Guide, Version 9.2.0. German Aerospace Center (DLR), Wessling (2018). DLR Report DLR-IB 565-01/2018
104. Chavez, P.S.: An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sens. Environ. **24**, 459–479 (1988)
105. Kaufman, Y.J., Wald, A.E., Remer, L.A., Gao, B.C., Li, R.R., Flynn, L.: The MODIS 2.1-μm channel – Correlation with visible reflectance for use in remote sensing of aerosol. IEEE Trans. Geosci. Remote Sens. **35**, 1286–1298 (1997)
106. Rahman, H., Dedieu, G.: SMAC: A simplified method for the atmospheric correction of satellite measurements in the solar spectrum. Int. J. Remote Sens. **15**(1), 123–143 (1994)
107. Song, C., Woodcock, C.E.: Monitoring forest succession with multitemporal Landsat images: Factors of uncertainty. IEEE Trans. Geosci. Remote Sens. **41**, 2557–2567 (2003)
108. Song, C., Woodcock, C.E., Seto, K.C., Lenney, M.P., Macomber, S.A.: Classification and change detection using Landsat TM data: When and how to correct atmospheric effects? Remote Sens. Environ. **75**, 230–244 (2001)
109. Chavez, P.S.: Radiometric calibration of Landsat thematic mapper multispectral images. Photogramm. Eng. Remote Sens. **55**, 1285–1294 (1989)
110. Kneizys, F.X., Shettle, E.P., Gallery, W.O., Chetwynd, J.H., Anderson, G.P., Abreu, L.W., Gallery, W.O., Selby, J.E.A., Clough, S.A.: User Guide to LOWTRAN 7. Environmental Research Papers No. 1010, p. 137 (1988)
111. Kneizys, F.X., Robertson, D.C., Abreu, L.W., Acharya, P., Anderson, G.P., Rothman, L.S., Chetwynd, J.H., Selby, J.E.A., Shettle, E.P., Gallery, W.O., Berk, A., Clough, S.A., Bernstein, L.S.: The MODTRAN 2/3 Report and LOWTRAN 7 Model (1996). www.gps.caltech.edu/~vijay/pdf/modrept.pdf, Accessed 30 Jan 2019
112. Berk, A., Anderson, G. P., Bernstein, L. S., Acharya, P. K., Dothe, H., Matthew, M. W. ,Adler-Golden, S. M., Chetwynd, Jr., J. H., Richtsmeier, S. C., Pukall, B., Allred, C. L., Jeong, L. S., Hoke, M. L.: MODTRAN4 Radiative Transfer Modeling for Atmospheric Correction. SPIE Proceeding, Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research III, Volume 3756, (1999), https://www.spectral.com/wp-content/uploads/2017/04/MODTRAN4_Radiative_Transfer.pdf (last access: 2019/02/10)
113. Berk, A., Anderson, G.P., Acharya, P.K., Chetwynd, Jr., J. H.,Bernstein, L.S., Dothe, H., Matthew, M.W., Adler-Golden, S.M., Richtsmeier, S. C., Pukall, B., Allred, C. L., Jeong, L. S., Hoke, M. L.: MODTRAN4 User's Manual. Air Force Research Laboratory, Space Vehicles Directorate, Air Force Materiel Command, Hanscom AFB, MA 01731-3010.- p. 94. (1999), ftp://ftp.pmodwrc.ch/pub/Vorlesung%20K+S/MOD4_user_guide.pdf (last access: 2019/02/10)
114. Popp, T.: Correcting atmospheric masking to retrieve the spectral albedo of land surfaces from satellite. Int. J. Remote Sens. **16**, 3483–3508 (1995)
115. Fraser, R.S., Ferrare, R.A., Kaufman, Y.J., Mattoo, S.: Algorithm for atmospheric corrections of aircraft and satellite imagery. Int. J. Remote Sens. **13**, 541–557 (1992)
116. Richter, R., Schlaepfer, D., and Mueller, A.: Operational atmospheric correction for imaging spectrometers accounting the smile effect. IEEE TGRS, Vol. 49, 1772–1780 (2011)
117. Richter, R.: Atmospheric correction of satellite data with haze removal including a haze/clear transition region. Comput. Geosci. **22**, 675–681 (1996)
118. Richter, R.: Correction of atmospheric and topographic effects for high spatial resolution satellite imagery. Int. J. Remote Sens. **18**, 1099–1111 (1997)
119. Kaufman, Y.J., Sendra, C.: Automatic atmospheric correction. Int. J. Remote Sens. **9**, 1357–1381 (1988)
120. Maiersperger, T.K., Scaramuzza, P.L., Leigh, L., Shrestha, S., Gallo, K.P., Jenkerson, C.B., Dwyer, J.L.: Characterizing LEDAPS surface reflectance products by comparisons with AERONET, field spectrometer, and MODIS data. Remote Sens. Environ. **136**, 1–13 (2013)
121. Claverie, M., Vermote, E.F., Franch, B., Masek, J.G.: Evaluation of the Landsat-5 TM and Landsat-7 ETM+ surface reflectance products. Remote Sens. Environ. **169**, 390–403 (2015)
122. Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Müller-Wilm, U., Gascon, F.: Sen2Cor version 2.5: Radiometric Validation – 2nd Sentinel-2 Validation Teem Meeting 29–31 January 2018. ESA-ESRIN, Frascati, p. 14 (2018)
123. Louis, J., Pflug, B., Main-Knorn, M., Debaecker, V., Müller-Wilm, U., Gascon, F.: Integration and Assimilation of metrological (ECMWF) aerosol estimates into Sen2Cor atmospheric correction. In: Proc. IGARSS, p. 4. (2018)
124. Wald, L.: Some terms of reference in data fusion. IEEE Trans. Geosci. Remote Sens. **37**(3), 1190–1193 (1999)
125. Pohl, C., van Genderen, J.L.: Multisensor image fusion in remote sensing: Concepts, methods and applications. Int. J. Remote Sens. **19**(5), 823–854 (1998)

126. Smith, A.R.: Color gamut transformation pairs. In: Proc. 5th Annu. Conf. Comput. Graph. Interact. Tech. (SIGGRAPH '78), pp. 12–19. (1978)
127. Joblove, G.H., Greenberg, D.: Color spaces for computer graphics. ACM SIGGRAPH Comput. Graph. **12**(3), 20–25 (1978)
128. Hayden, R., Henkel, G.W.J., Bare, J.E.: Application of the IHS color transform to the processing of multisensor data and image enhancement. In: Proc. Int. Symp. Remote Sens. Environ. pp. 599–616. (1982)
129. Polidori, L., Mangolini, M.: Potentials and limitations of multisensor data fusion. In: Proc. EARSeL Conf. Fus. Earth Data, pp. 13–19. (1996)
130. Cetin, M., Mussaoglu, N.: Merging hyperspectral and panchromatic image data: Qualitative and quantitative analysis. Int. J. Remote Sens. **30**(7), 1779–1804 (2009)
131. Vivone, G., Alparone, L., Chanussot, J., Mura, M.D., Garzelli, A., Licciardi, G.A., Restaino, R., Wald, L.: A critical comparison among pansharpening algorithms. IEEE Trans. Geosci. Remote Sens. **53**(5), 2565–2586 (2015)
132. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Process. Lett. **9**(3), 81–84 (2002)
133. Goodman, I.R., Mahler, R.P., Nguyen, H.T.: Mathematics of Data Fusion. Theory and Decision Library B, vol. 37. Springer, Dordrecht (1997)
134. Haberäcker, P.: Digitale Bildverarbeitung – Grundlagen und Anwendungen. Hanser, München, p. 404 (1989)
135. Canty, M.J.: Image Analysis, Classification and Change Detection in Remote Sensing – With Algorithms for ENVI/IDL and Python. CRC, Boca Raton (2014)
136. Pavlides, T.: Algorithmen zur Grafik und Bildverarbeitung. Heise, Hannover, p. 508 (1990)
137. Bässmann, H., Kreyss, J.: Bildverarbeitung ad Oculos, 3rd edn. Springer, Berlin, p. 473 (1998)
138. Liedke, C.-E., Ender, M.: Wissensbasierte Bildverarbeitung. Nachrichtentechnik, vol. 19. Springer, Berlin, Heidelberg (1989)
139. Rees, W.G.: Physical Principles of Remote Sensing. Cambridge Univ. Press, Cambridge, p. 345 (2001)
140. Lu, D., Weng, Q.: Review article – A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens. **28**(5), 823–870 (2007)
141. Prasad, S.V.S., Savithri, T.S., Krishn, I.V.M.: Techniques in image classification. A survey. Version 1. Glob. J. Res. Eng. F **15**(6), 17–32 (2015)
142. Borg, E., Fichtelmann, B., Asche, H.: Cloud classification in JPEG-compressed remote sensing data (LANDSAT 7/ETM+). Lect. Notes Comput. Sci. **7334**, 347–357 (2012)
143. Zhukov, B., Oertel, D., Lanzl, F., Reinhäckel, G.: Unmixing-based multisensor multiresolution image fusion. IEEE Trans. Geosci. Remote Sens. **37**(3), 1212–1226 (1999)
144. Hansen, M., Dubayah, R., Defries, R.: Classification trees: An alternative to traditional land cover classifiers. Int. J. Remote Sens. **17**(5), 1075–1081 (1996)
145. Schowengerdt, R.A.: On the estimation of spatial-spectral mixing with classifier likelihood functions. Pattern Recognit. Lett. **17**(13), 1379–1387 (1996)
146. Pal, M., Mather, P.M.: An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sens. Environ. **86**, 554–565 (2003)
147. Chen, K.S., Tzeng, Y.C., Chen, C.F., Kao, W.L.: Land-cover classification of multispectral imagery using a dynamic learning neural network. Photogramm. Eng. Remote Sens. **61**, 403–408 (1995)
148. Witten, I.H., Frank, E., Hall, M.A.: Data Mining. Elsevier, Amsterdam, p. 629 (2011)
149. Hastie, T., Tibshirani, R., Friedman, J.H.: The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Springer, New York, p. 745 (2009)
150. Mitra, P., Shankar, B.U., Pal, S.K.: Segmentation of multispectral remote sensing images using active support vector machines. Pattern Recognit. Lett. **25**, 1067–1074 (2004)
151. Lohmann, G.: An evidential reasoning approach to the classification of satellite images. Lect. Notes Comput. Sci. **548**, 227–231 (1991)
152. Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M.: Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J. Photogramm. Remote Sens. **58**(3/4), 239–258 (2004)
153. Stuckens, J., Coppin, P.R., Bauer, M.E.: Integrating contextual information with per-pixel classification for improved land cover classification. Remote Sens. Environ. **71**, 282–296 (2000)
154. Congalton, R.G.: A review of assessing the accuracy of classification of remotely sensed data. Remote Sens. Environ. **37**, 35–46 (1991)
155. Congalton, R.G., Green, K.: Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, 2nd edn. CRC, Boca Raton, p. 192 (1999)
156. Dorigo, W.A., Zurita-Milla, R., De Wit, A.J.W., Brazile, J., Singh, R., Schaepman, M.E.: A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. Int. J. Appl. Earth Obs. Geoinf. **9**, 165–193 (2007)
157. Verrelst, J., Malenovský, Z., Van der Tol, C., Camps Valls, G., Gastellu Etchegorry, J.-P., Lewis, P., North, P., Moreno, J.: Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods. Surv. Geophys. **5**(1), 589–629 (2018)
158. Verhoef, W.: Theory of radiative transfer models applied in optical remote sensing of vegetation canopies, Thesis Landbouwuniversiteit Wageningen, pp. 310. Grafisch Service Centrum Van Gils b.v., Wageningen (1998)
159. Verhoef, W., Bach, H.: Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models. Remote Sens. Environ. **87**, 23–41 (2003)
160. Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., Wang, L.: Retrieval of canopy biophysical variables from bidirectional reflectance. Using prior information to solve the ill-posed inverse problem. Remote Sens. Environ. **84**, 1–15 (2002)
161. Garabedian, P.: Partial differential equations. Wiley, New York, p. 672 (1964)
162. Ertel, W.: Grundkurs Künstliche Intelligenz – Folien zum Buch (2018). www.hs-weingarten.de/~ertel/kibuch/kibuch-folien.pdf, Accessed 30 Jan 2019
163. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016). http://www.deeplearningbook.org/ (last access: 23 January 2019)
164. Bengio, Y.: Deep Learning (2015). http://www.iro.umontreal.ca/~bengioy/talks/lisbon-mlss-19juillet2015.pdf, Accessed 25 Jan 2019
165. Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A.: Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177 (2019)
166. Mitchell, T.: Machine Learning. McGraw-Hill, New York, p. 414 (1997)
167. Boutaba, R., Salahuddin, M.A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., Caicedo, O.M.: A comprehensive survey on machine learning for networking: Evolution, applications and research opportunities. J. Internet Serv. Appl. **9**, 16 (2018)

168. Solares, C., Sanz, A.M.: Bayesian network classifiers. An application to remote sensing image classification. In: Proc. 6th WSEAS Int. Conf. Neural Networks, pp. 62–67. (2005)
169. Solares, C., Sanz, A.M.: Different Bayesian network models in the classification of remote sensing images. Lect. Notes Comput. Sci. **4881**, 10–16 (2007)
170. Mountrakis, G., Ogole, J.C.: Support vector machines in remote sensing: A review. ISPRS J. Photogramm. Remote Sens. **66**(3), 247–259 (2011)
171. Belgiu, M., Drăguţ, L.: Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. **114**, 24–31 (2016)
172. Kolesnikov, A.: Efficient Algorithms for Vectorization and Polygonal Approximation. Dissertation. Univ. Joensuu, Joensuu, p. 110 (2003)
173. Borg, E., Ressl, R., Günther, K.: Fernerkundungsdaten und -produkte für die Landschaftsstrukturanalyse. In: Walz, U., Lutze, G., Schultz, A., Syrbe, R.-U. (eds.) Landschaftsstruktur – Methoden und Anwendungen im Kontext von naturräumlicher Vorprägung und Nutzung IÖR-Schriften, vol. 43, pp. 79–90. Sächsische Druck- und Verlagshaus AG, Dresden (2004)
174. H. Leser: Landschaftsökologie. Ansatz, Modelle, Methodik, Anwendung. 4. neu bearbeitete Auflage. Ulmer, Stuttgart 1997, ISBN 3-8252-0521-5, S. 145, 148 (UTB 521)
175. Tomlinson, C.J., Chapman, L., Thornes, J.E., Baker, C.: Remote sensing land surface temperature for meteorology and climatology: A review. Meteorol. Appl. **18**, 296–306 (2011)
176. Adler, R.F., Negri, A.J.: A satellite infrared technique to estimate tropical convective and stratiform rainfall. J. Appl. Meteorol. **27**, 30–51 (1988)
177. Vicente, G.A., Scofield, R.A., Menzel, W.P.: The operational GOES infrared rainfall estimation technique. Bull. Am. Meteorol. Soc. **79**, 1883–1898 (1998)
178. Velden, C.S., Olander, T.L., Wanzong, S.: The impact of multispectral GOES-8 wind information on Atlantic tropical cyclone track forecasts in 1995. Part I: Dataset methodology, description, and case analysis. Mon. Weather Rev. **126**, 1202–1218 (1998)
179. Velden, C.S., Stettner, D., Daniels, J.: Wind vector fields derived from GOES rapid-scan imagery. In: Proc. 10th Conf. Satell. Meteorol. Oceanogr. pp. 20–23. (2000)
180. King, M.D., Menzel, W.P., Kaufman, Y.J., Tanré, D., Gao, B.C., Platnick, S., Ackerman, S.A., Remer, L., Pincus, R., Hubanks, P.A.: Cloud and aerosol properties, precipitable water, and profiles of temperature and humidity from MODIS. IEEE Trans. Geosci. Remote Sens. **41**(2), 442–458 (2003)
181. Morsy, M., Dietrich, P., Scholten, T., Michaelides, S., Borg, E., Sherief, Y.: The potential of using satellite-related precipitation data sources in arid regions (CHAPTER 7). In: Michaelides, S. (ed.) Precipitation Science - Measurement, Remote Sensing, Microphysics, and Modeling, pp. 201–237. Elsevier Inc., Amsterdam (2021). Available from https://doi.org/10.1016/B978-0-12-822973-6.00002-0.
182. Saunders, R.W.: An automated scheme for the removal of cloud contamination from AVHRR radiances over Western Europe. Int. J. Remote Sens. **7**(7), 867–886 (1986)
183. Saunders, R.W., Kriebel, K.T.: An improved method for detecting clear sky and cloudy radiances from AVHRR data. Int. J. Remote Sens. **9**(1), 123–150 (1988)
184. Feijt, A., De Valk, P., Van Der Veen, S.: Cloud detection using meteosat imagery and numerical weather prediction model data. J. Appl. Meteorol. Clim. **39**(7), 1017–1030 (2010)
185. Taravat, A., Proud, S., Peronaci, S., Del Frate, F., Oppelt, N.: Multilayer perceptron neural networks model for meteosat second generation SEVIRI daytime cloud masking. Remote Sens. **7**(2), 1529–1539 (2015)
186. Szturc, J., Osrodka, K., Jurczyk, A., Otop, I., Linkowska, J., Bochenek, B., Pasierb, M. : Quality control and verification of precipitation observations, estimates, and forecasts (CHAPTER 3). In: Michaelides, S. (ed.) Precipitation Science - Measurement, Remote Sensing, Microphysics, and Modeling, pp. 91–133. Elsevier Inc., Amsterdam (2021). Available from https://doi.org/10.1016/B978-0-12-822973-6.00002-0
187. WMO: Weather Radar – Part 1: System Performance and Operation (Draft Text of the Common ISO/WMO Standard) (2008). http://www.wmo.int/pages/prog/www/IMOP/ISO/Std_Radar_Part1_For_Members_Review.pdf, Accessed 23 Nov 2018
188. Marshall, J.S., Palmer, W.M.: The distribution of raindrops with size. J. Atmos. Sci. **5**, 165–166 (1948)
189. Atlas, D., Srivastava, R.C., Sekhon, R.S.: Doppler radar characteristics of precipitation at vertical incidendence. Rev. Geophys. Space Phys. **11**, 1–35 (1973)
190. Kumar, L.S., Lee, Y.H., Yeo, J.X., Ong, J.T.: Tropical rain classification and estimation of rain from Z-R (reflectivity-rain rate) relationships. Prog. Electromagn. Res. B **32**, 107–127 (2011)
191. Lehner, S., Tings, B.: Maritime products using TerraSAR-X and Sentinel-1 imagery. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. **XL**(7/W3), 967–973 (2015)
192. Gayathri, K.D., Ganasri, B.P., Dwarakish, G.S.: Applications of remote sensing in satellite oceanography: A review, International Conference on Water Resources, Coastal and Ocean Engineering (ICWRCOE'15). Aquat. Procedia **4**, 579–584 (2015)
193. Klemas, V.: Fisheries applications of remote sensing: An overview. Fish. Res. **148**, 124–136 (2013)
194. Fingas, M., Brown, C.E.: A review of oil spill remote sensing. Sensors **18**(1), 91 (2018)
195. Maximenko, N., Corradi, P., Law, K.L., Van Sebille, E., Garaba, S.P., Lampitt, R.S., Galgani, F., Martinez-Vicente, V., Goddijn-Murphy, L., Veiga, J.M., Thompson, R.C., Maes, C., Moller, D., Löscher, C.R., Addamo, A.M., Lamson, M.R., Centurioni, L.R., Posth, N.R., Lumpkin, R., Vinci, M., Martins, A.M., Pieper, C.D., Isobe, A., Hanke, G., Edwards, M., Chubarenko, I.P., Rodriguez, E., Aliani, S., Arias, M., Asner, G.P., Brosich, A., Carlton, J.T., Chao, Y., Cook, A.-M., Cundy, A.B., Galloway, T.S., Giorgetti, A., Goni, G.J., Guichoux, Y., Haram, L.E., Hardesty, B.D., Holdsworth, N., Lebreton, L., Leslie, H.A., Macadam-Somer, I., Mace, T., Manuel, M., Marsh, R., Martinez, E., Mayor, D.J., Le Moigne, M., Molina Jack, M.E., Mowlem, M.C., Obbard, R.W., Pabortsava, K., Robberson, B., Rotaru, A.-E., Ruiz, G.M., Spedicato, M.T., Thiel, M., Turra, A., Wilcox, C.: Toward the Integrated Marine Debris Observing System. Front. Mar. Sci. **6**, 447 (2019) https://doi.org/10.3389/fmars.2019.00447
196. Kako, S., Isobe, A., Magome, S.: Sequential monitoring of beach litter using webcams. Mar. Pollut. Bull. **60**(5), 775–779 (2010)
197. Kako, S., Isobe, A., Magome, S.: Low altitude remote-sensing method to monitor marine and beach litter of various colors using a balloon equipped with a digital camera. Mar. Pollut. Bull. **64**(6), 1156–1162 (2012)
198. Odum, E.P.: Grundlagen der Ökologie. Thieme, Stuttgart, p. 836 (1983)
199. Gholizadeh, M.H., Melesse, A.M., Reddi, L.: A comprehensive review on water quality parameters estimation using remote sensing techniques. Sensors **16**(8), 43 (2016)
200. Kutser, T., Arst, H., Miller, T., Käärmann, L., Milius, A.: Telespectrometrical estimation of water transparency, chlorophyll-a and total phosphorus concentration of Lake Peipsi. Int. J. Remote Sens. **16**(16), 3069–3085 (1995)
201. Gómez, B., Mertes, L.A.K., Phillpe, J.D., Magilligan, F.J., James, L.A.: Sediment characteristics of an extreme flood: 1993 upper Mississippi River valley. Geology **23**(11), 963–966 (1995)

202. Mannheim, S., Segl, K., Heim, B., Kaufmann, H.: Monitoring of lake water quality using hyperspectral CHRIS-PROBA data. In: Proc. 2nd CHRIS/PROBA Workshop, pp. 28–30. (2004)
203. Dekker, A.G., Vos, R., Peters, S.: Analytical algorithms for lake water TSM estimation for retrospective analyses of TM and SPOT sensor data. Int. J. Remote Sens. **23**(1), 15–35 (2002)
204. Wloczyk, C., Richter, R., Borg, E., Neubert, W.: Sea and lake surface temperaure retrieval from Landsat thermal data in Northern Germany. Int. J. Remote Sens. **27**(12), 2489–2502 (2006)
205. Markogianni, V., Dimitriou, E., Tzortziou, M.: Monitoring of chlorophyll-a and turbidity in Evros River (Greece) using Landsat imagery. In: Proc. 1st Int. Conf. Remote Sens. Geoinf. Environ. (RSCy2013) (2013)
206. Duan, W., Takara, K., He, B., Luo, P., Nover, D., Yamashiki, Y.: Spatial and temporal trends in estimates of nutrient and suspended sediment loads in the Ishikari River, Japan, 1985 to 2010. Sci. Total Environ. **461/462**, 499–508 (2013)
207. Karaska, M.A., Huguenin, R.L., Beacham, J.L., Wang, M.-H., Jensen, J.R., Kaufmann, R.S.: AVIRIS measurements of chlorophyll, suspended minerals, dissolved organic carbon, and turbidity in the Neuse River, North Carolina. Photogramm. Eng. Remote Sens. **70**(1), 125–133 (2004)
208. Jensen, J.R., Hodgson, M.E., Christensen, E., Mackey Jr, H.E., Tinney, L.R.: Remote sensing inland wetlands: A multispectral approach. Photogramm. Eng. Remote Sens. **52**(1), 87–100 (1986)
209. Milton, E.J., Hughes, P.D., Anderson, K., Schulz, J., Lindsay, R., Kelday, S.B., Hill, C.T.: Remote Sensing of Bog Surfaces (2005) JNCC Report No. 366, pp. 99
210. Dabrowska-Zielinska, K., Budzynska, M., Tomaszewska, M., Bartold, M., Gatkowska, M., Malek, I., Turlej, K., Napiorkowska, M.: Monitoring wetlands ecosystems using ALOS PALSAR (L-Band, HV) supplemented by optical data: A case study of Biebrza Wetlands in northeast Poland. Remote Sens. **6**(2), 1605–1633 (2014)
211. Schneider, S.: Luftbild und Luftbildinterpretation. Walter de Gruyter, Berlin, p. 519 (1974)
212. Forkuo, E.K.: Digital terrain modeling in a GIS environment. ISPRS Archives **37**(B2), 1023–1030 (2008)
213. Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R.: The use of remote sensing in soil and terrain mapping – A review. Geoderma **162**, 1–19 (2011)
214. Castaldi, F., Chabrillat, S., Jones, A., Vreys, K., Bomans, B., Van Wesemael, B.: Soil organic carbon estimation in croplands by hyperspectral remote APEX data using the LUCAS topsoil database. Remote Sens. **10**(2), 153 (2018)
215. Chabrillat, S., Eisele, A., Guillaso, S., Rogaß, C., Ben-Dor, E., Kaufmann, H.: HYSOMA: An easy-to-use software interface for soil mapping applications of hyperspectral imagery. In: Proc. EARSeL SIG Imaging Spectrosc. Workshop, p. 7. (2011)
216. van der Meer, F.D., van der Werff, H.M.A., van Ruitenbeek, F.J.A.: Potential of ESA's Sentinel-2 for geological applications. Remote Sens. Environ. **148**, 124–133 (2014)
217. Mielke, C., Boesche, N.K., Rogass, C., Segl, K., Gauert, C., Kaufmann, H.: Potential applications of the Sentinel-2 multispectral sensor and the EnMAP hyperspectral sensor in mineral exploration. EARSeL EProceedings **13**(2), 93–102 (2014)
218. Driscoll, R.S.: Color Aerial Photography – A New View for Range Management (1971) USDA Forest Service Research Paper RM-67
219. Cernansky, R.: Biodiversity moves beyond counting species. Nature **546**, 22–24 (2017)
220. Lausch, A., Bannehr, L., Beckmann, M., Boehm, C., Feilhauer, H., Hacker, J.M., Heurich, M., Jung, A., Klenke, R., Neumann, C., Pause, M., Rocchini, D., Schaepman, M.E., Schmidtlein, S., Schulz, K., Selsam, P., Settele, J., Skidmore, A.K., Cord, A.F.: Linking Earth observation and taxonomic, structural and functional biodiversity: Local to ecosystem perspectives. Ecol. Indic. **70**, 317–339 (2016)
221. Lausch, A., Erasmi, S., King, D., Magdon, P., Heurich, M.: Understanding forest health with remote sensing – Part II: A review of approaches and data models. Remote Sens. **9**, 129 (2017)
222. Lausch, A., Erasmi, S., King, D.J., Magdon, P., Heurich, M.: Understanding forest health with remote sensing – Part I: A review of spectral traits, processes and remote-sensing characteristics. Remote Sens. **8**(12), 1029 (2016)
223. Lausch, A., Bastian, O., Klotz, S., Leitão, P.J., Jung, A., Rocchini, D., Schaepman, M.E., Skidmore, A.K., Tischendorf, L., Knapp, S.: Understanding and assessing vegetation health by in-situ species and remote sensing approaches. Methods Ecol. Evol. **9**(8), 1799–1809 (2018)
224. Asner, G.P., Martin, R.E.: Airborne spectranomics: Mapping canopy chemical and taxonomic diversity in tropical forests. Front. Ecol. Environ. **7**, 269–276 (2009)
225. Asner, G.P., Martin, R.E.: Spectranomics: Emerging science and conservation opportunities at the interface of biodiversity and remote sensing. Glob. Ecol. Conserv. **8**, 212–219 (2016)
226. Schweiger, A.K., Cavender-Bares, J., Townsend, P.A., Hobbie, S.E., Madritch, M.D., Wang, R., Tilman, D., Gamon, J.A.: Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. Nat. Ecol. Evol. **2**, 976–982 (2018)
227. McManus, K.M., Asner, G.P., Martin, R.E., Dexter, K.G., Kress, W.J., Field, C.B.: Phylogenetic structure of foliar spectral traits in tropical forest canopies. Remote Sens. **8**, 1–16 (2016)
228. Richter, R., Reu, B., Wirth, C., Doktor, D.: The use of airborne hyperspectral data for tree species classification in a species-rich Central European forest area. Int. J. Appl. Earth Obs. Geoinf. **52**, 464–474 (2016)
229. Tanase, M.A., Kennedy, R., Aponte, C.: Radar burn ratio for fire severity estimation at canopy level: An example for temperate forests. Remote Sens. Environ. **170**, 14–31 (2015)
230. Heurich, M., Thoma, F.: Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (*Fagus sylvatica*) and Norway spruce (*Picea abies*) forests. Forestry **81**, 645–661 (2008)
231. Schneider, F.D., Morsdorf, F., Schmid, B., Petchey, O.L., Hueni, A., Schimel, D.S., Schaepman, M.E.: Mapping functional diversity from remotely sensed morphological and physiological forest traits. Nat. Commun. **8**, 1441 (2017)
232. Lausch, A., Borg, E., Bumberger, J., Dietrich, P., Heurich, M., Huth, A., Jung, A., Klenke, R., Knapp, S., Mollenhauer, H., Paasche, H., Paulheim, H., Pause, M., Schweitzer, C., Schmullius, C., Settele, J., Skidmore, A., Wegmann, M., Zacharias, S., Kirsten, T., Schaepman, M.: Understanding forest health with remote sensing – Part III: Requirements for a scalable multi-source forest health monitoring network based on data science approaches. Remote Sens. **10**, 1120 (2018)
233. Bardgett, R.D., Wardle, D.A.: Aboveground-belowground linkages: Biotic interactions, ecosystem processes, and global change. Oxford Univ. Press, Oxford (2010)
234. Roser, M., Ritchie, H.: Yields and Land Use in Agriculture (2018) https://ourworldindata.org/yields-and-land-use-in-agriculture, Accessed 28 Nov 2018
235. Wulder, M.A., Coops, N.C.: Make Earth observations open access. Nature **513**, 30–31 (2014)
236. Griffiths, P., Nendel, C., Hostert, P.: Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. Remote Sens. Environ. **220**, 135–151 (2019)
237. Blume, H.-P., Brümmer, G.W., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretschmar, R., Stahr, K., Wilke, B.-M.: Lehrbuch der Bodenkunde. Springer, Berlin, Heidelberg, p. 569 (2016)

238. Buringh, P.: The applications of aerial photographs in soil surveys. In: Manual of Photographic Interpretation, pp. 633–666. ASPRS, Washington (1960)
239. Wloczyk, C., Borg, E., Richter, R., Miegel, K.: Estimation of instantaneous air temperature above vegetation and soil surfaces from Landsat 7 ETM+ data in Northern Germany. Int. J. Remote Sens. **32**(24), 9119–9136 (2011)
240. Ben-Dor, E., Inbar, Y., Chen, Y.: The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. Remote Sens. Environ. **61**(1), 1–15 (1997)
241. Gerighausen, H., Menz, G., Kaufmann, H.: Spatially explicit estimation of clay and organic carbon content in agricultural soils using multi-annual imaging spectroscopy data. Appl. Environ. Soil Sci. (2012) https://doi.org/10.1155/2012/868090
242. Mulder, V.L., de Bruin, S., Weyermann, J., Kokaly, R.F., Schaepman, M.E.: Characterizing regional soil mineral composition using spectroscopy and geostatistics. Remote Sens. Environ. **139**, 415–429 (2013)
243. Wulf, H., Mulder, V.L., Schaepman, M.E., Keller, A., Jörg, A., Claudio, P.: Remote Sensing of Soils. Univ. Zurich, Zurich (2014) https://doi.org/10.13140/2.1.1098.0649. Technical Report
244. Franke, J., Keuck, V., Siegert, F.: Assessment of grassland use intensity by remote sensing to support conservation schemes. J. Nat. Conserv. **20**, 125–134 (2012)
245. Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P.H., Estel, S., Haberl, H., Hostert, P., Jepsen, M.R., Kastner, T., Levers, C., Lindner, M., Plutzar, C., Verkerk, P.J., van der Zanden, E.H., Reenberg, A.: Challenges and opportunities in mapping land use intensity globally. Curr. Opin. Environ. Sustain. **5**(5), 484–493 (2013)
246. Graesser, J., Aide, T.M., Grau, H.R., Ramankutty, N.: Cropland/-pastureland dynamics and the slowdown of deforestation in Latin America. Environ. Res. Lett. **10**(3), 34017 (2015)
247. Wachendorf, M., Fricke, T., Möckel, T.: Remote sensing as a tool to assess botanical composition, structure, quantity and quality of temperate grasslands. Grass Forage Sci. **73**(1), 1–14 (2018)
248. Mahlein, A.-K.: Plant disease detection by imaging sensors – Parallels and specific demands for precision agriculture and plant phenotyping. Plant Dis. **100**(2), 241–251 (2016)
249. Bill, R.: Interoperable GIS-Infrastruktur im landwirtschaftlichen Betrieb. In: Heineke, H.J. (ed.) Geographische Informationssysteme in der Landwirtschaft und im ländlichen Raum: Defizite und Entwicklungspotentiale KTBL-Schriften, vol. 428, pp. 69–74. KTBL-Schriften-Vertrieb, Münster (2004)
250. Thenkabail, P.S., Enclona, E.A., Ashton, M.S., Van Der Meer, B.: Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. Remote Sens. Environ. **91**, 354–376 (2004)
251. Ríos-Casanova, L., Valiente-Banuet, A., Rico-Gray, V.: Ant diversity and its relationship with vegetation and soil factors in an alluvial fan of the Tehuacán Valley, Mexico. Acta Oecol. **29**, 316–323 (2006)
252. Orwin, K.H., Buckland, S.M., Johnson, D., Turner, B.L., Smart, S., Oakley, S., Bardgett, R.D.: Linkages of plant traits to soil properties and the functioning of temperate grassland. J. Ecol. **98**, 1074–1083 (2010)
253. Mueller, K.E., Hobbie, S.E., Chorover, J., Reich, P.B., Hale, C.M., Jagodzin, A.M.: Effects of litter traits, soil biota, and soil chemistry on soil carbon stocks at a common garden with 14 tree species. Biogeochemistry **123**, 313–327 (2015)
254. Krishnan, P., Kochendorfer, J., Dumas, E.J., Guillevic, P.C., Baker, C.B., Meyers, T.P., Martos, B.: Comparison of in-situ, aircraft, and satellite land surface temperature measurements over a NOAA Climate Reference Network site. Remote Sens. Environ. **165**, 249–264 (2015)
255. de Vries, F.T., Manning, P., Tallowin, J.R.B., Mortimer, S.R., Pilgrim, E.S., Harrison, K.A., Hobbs, P.J., Quirk, H., Shipley, B., Cornelissen, J.H.C., Kattge, J., Bardgett, R.D.: Abiotic drivers and plant traits explain landscape-scale patterns in soil microbial communities. Ecol. Lett. **15**(11), 1230–1239 (2012)
256. Atkinson, N.J., Urwin, P.E.: The interaction of plant biotic and abiotic stresses: From genes to the field. J. Exp. Bot. **63**(10), 3523–3543 (2012)
257. Hall, K., Lamont, N.: Zoogeomorphology in the Alpine: Some observations on abiotic–biotic interactions. Geomorphology **55**, 219–234 (2003)
258. Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Harold, A., Mooney, M.O., Poff, N.L., Sykes, M.T., Walker, B.H., Walker, M., Wall, D.H.: Biodiversity: Global biodiversity scenarios for the year 2100. Sci. New Ser. **287**(5459), 1770–1775 (2005)
259. Beier, P., Brost, B.: Uso de elementos territoriales para planificar para el cambio climático: Conservando las arenas, no los actors. Conserv. Biol. **24**, 701–710 (2010)
260. Pressey, R.L., Cabeza, M., Watts, M.E., Cowling, R.M., Wilson, K.A.: Conservation planning in a changing world. Trends Ecol. Evol. **22**(11), 583–592 (2007)
261. Dörnhöfer, K., Oppelt, N.: Remote sensing for lake research and monitoring – Recent advances. Ecol. Indic. **64**, 105–122 (2016)
262. Wulder, M.A., White, J.C., Loveland, T.R., Woodcock, C.E., Belward, A.S., Cohen, W.B., Fosnight, E.A., Shaw, J., Masek, J.G., Roy, D.P.: The global Landsat archive: Status, consolidation, and direction. Remote Sens. Environ. **185**, 271–283 (2016)
263. Braun, G.: Vegetationsgeographische Untersuchungen im NW-Karakorum (Pakistan): Kartierung der aktuellen Vegetation und Rekonstruktion der potentiellen Waldverbreitung auf der Basis von Satellitendaten, Gelände- und Einstrahlungsmodellen. Bonner geographische Abhandlungen, vol. 7643. Dümmler, Bonn, p. 156 (1996)
264. Jones, T.G., Coops, N.C., Sharma, T.: Assessing the utility of airborne hyperspectral and LIDAR data for species distribution mapping in the coastal Pacific Northwest, Canada. Remote Sens. Environ. **114**, 2841–2852 (2010)
265. Sollins, P.: Factors influencing species composition in tropical lowland rain forest: Does soil matter? Ecology **79**, 23–30 (1998)
266. Rowan, L.C., Mars, J.C.: Lithologic mapping in the Mountain Pass, California area using advanced spaceborne thermal emission and reflection radiometer (ASTER) data. Remote Sens. Environ. **84**, 350–366 (2003)
267. Gholizadeh, H., Gamon, J.A., Zygielbaum, A.I., Wang, R., Schweiger, A.K., Cavender-Bares, J.: Remote sensing of biodiversity: Soil correction and data dimension reduction methods improve assessment of α-diversity (species richness) in prairie ecosystems. Remote Sens. Environ. **206**, 240–253 (2018)
268. Anderson, K., Croft, H.: Remote sensing of soil surface properties. Prog. Phys. Geogr. **33**(4), 457–473 (2009)
269. McColl, K.A., Alemohammad, S.H., Akbar, R., Konings, A.G., Yueh, S., Entekhabi, D.: The global distribution and dynamics of surface soil moisture. Nat. Geosci. **10**, 100–104 (2017)
270. Vereecken, H., Schnepf, A., Hopmans, J.W., Javaux, M., Or, D., Roose, T., Vanderborght, J., Young, M.H., Amelung, W., Aitkenhead, M., Allison, S.D., Assouline, S., Baveye, P., Berli, M., Brüggemann, N., Finke, P., Flury, M., Gaiser, T., Govers, G., Ghezzehei, T., Hallett, P., Hendricks Franssen, H.J., Heppell, J., Horn, R., Huisman, J.A., Jacques, D., Jonard, F., Kollet, S., Lafolie, F., Lamorski, K., Leitner, D., McBratney, A., Minasny, B., Montzka, C., Nowak, W., Pachepsky, Y., Padarian, J., Romano,

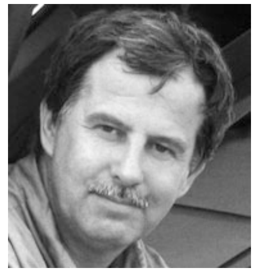
N., Roth, K., Rothfuss, Y., Rowe, E.C.C., Schwen, A., Šimůnek, J., Tiktak, A., Van Dam, J., van der Zee, S.E.A.T.M., Vogel, H.J., Vrugt, J.A., Wöhling, T., Young, I.M.: Modeling soil processes: Review, key challenges, and new perspectives. Vadose Zone J. **15**(5), 57 (2016)
271. Vereecken, H., Huisman, J.A., Bogena, H., Vanderborght, J., Vrugt, J.A., Hopmans, J.W.: On the value of soil moisture measurements in vadose zone hydrology: A review. Water Resour. Res. **44**(4), W00D06 (2008)
272. Vogel, H.J., Bartke, S., Daedlow, K., Helming, K., Kögel-Knabner, I., Lang, B., Rabot, E., Russell, D., Stößel, B., Weller, U., Wiesmeier, M., Wollschläger, U.: A systemic approach for modeling soil functions. Soil **4**, 83–92 (2018)
273. Rabot, E., Wiesmeier, M., Schlüter, S., Vogel, H.J.: Soil structure as an indicator of soil functions: A review. Geoderma **134**, 122–137 (2018)
274. Wessolek, G.: Bodenwasserhaushalt. In: Fohrer, N., Bormann, H., Miegel, K., Casper, M., Bronstert, A., Schumann, A., Weiler, M. (eds.) Hydrologie, pp. 69–90. Haupt, Bern (2016)
275. Ulaby, F.T., Dubois, P.C., van Zyl, J.: Radar mapping of surface soil moisture. J. Hydrol. **184**(1/2), 57–84 (1996)
276. Moreira, A., Prats, P., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K.A.: Tutorial on synthetic aperture radar. IEEE Geosci. Remote Sens. Mag. **1**(1), 6–43 (2013)
277. Jagdhuber, T., Member, S., Hajnsek, I., Member, S., Bronstert, A., Papathanassiou, K.P., Member, S.: Soil moisture estimation under low vegetation cover using a multi-angular polarimetric decomposition. IEEE Trans. Geosci. Remote Sens. **51**(4), 2201–2215 (2013)
278. Jagdhuber, T.: Soil Parameter Retrieval under Vegetation Cover Using SAR Polarimetry. Dissertation. Univ. Potsdam, Potsdam, p. 270 (2012)
279. Bruckler, L., Witono, H., Stengel, P.: Near surface soil moisture estimation from microwave measurements. Remote Sens. Environ. **26**(2), 101–121 (1988)
280. Zribi, M., Le Hégarat-Mascle, S., Ottlé, C., Kammoun, B., Guerin, C.: Surface soil moisture estimation from the synergistic use of the (multi-incidence and multi-resolution) active microwave ERS wind scatterometer and SAR data. Remote Sens. Environ. **86**(1), 30–41 (2003)
281. Pasolli, L., Notarnicola, C., Bertoldi, G., Bruzzone, L., Remelgado, R., Greifeneder, F., Niedrist, G., Della Chiesa, S., Tappeiner, U., Zebisch, M.: Estimation of soil moisture in mountain areas using SVR technique applied to multiscale active radar images at C-band. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **8**(1), 262–283 (2015)
282. Dobson, M.C., Ulaby, F.T.: Preliminary evaluation of the SIRB response to soil moisture, surface roughness, and crop canopy cover. IEEE Trans. Geosci. Remote Sens. **24**, 517–526 (1986)
283. Dubois, P.C., Engman, T.: Measuring soil moisture with imaging radars. IEEE Trans. Geosci. Remote Sens. **33**, 915–926 (1995)
284. Bindlish, R., Jackson, T.J., van der Velde, R.: High resolution soil moisture mapping using AIRSAR observations during SMEX02. In: IEEE Int. Conf. Geosci. Remote Sens. Symp. pp. 2324–2327. (2006) https://doi.org/10.1109/IGARSS.2006.601
285. Tsang, L., Kong, J.A., Ding, K.-H.: Scattering of Electromagnetic Waves. Theories and Applications. Wiley, New York, p. 445 (2000)
286. Elfouhaily, T. M., Guerin, C.-A.: A Critical Survey of Approximate Scattering Theories from Random Rough Surfaces. Wave Random Complex Media, **14**(4), 1–40 (2004) https://doi.org/10.1088/0959-7174/14/4/R01
287. Morrison, K., Bennett, J.C., Nolan, M., Menon, R.: Laboratory measurement of the DInSAR response to spatiotemporal variations in soil moisture. IEEE Trans. Geosci. Remote Sens. **49**(10), 3815–3823 (2011)
288. de Zan, F., Parizzi, A., Prats-Iraola, P., López-Dekker, P.: A SAR interferometric model for soil moisture. IEEE Trans. Geosci. Remote Sens. **52**(1), 418–425 (2014)
289. Zwieback, S., Hensley, S., Hajnsek, I.: Assessment of soil moisture effects on L-band radar interferometry. Remote Sens. Environ. **164**, 77–89 (2015)
290. Kerr, Y.H., Waldteufel, P., Wigneron, J.P., Martinuzzi, J.M., Font, J., Berger, M.: Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. IEEE Trans. Geosci. Remote Sens. **39**(8), 1729–1735 (2001)
291. Wigneron, J.-P.: Soil moisture retrieval algorithms in the framework of the SMOS mission: Current status and requirements for the EuroSTARRS Campaign. In: Fletcher, P. (ed.) Proc. First Results Workshop Eurostarrs, Wise, Losac Campaigns (2002), pp. 199–202. ESA Pub Division, Noordwijk (2003)
292. Huo, A., Zhang, J., Cheng, Y., Yi, X., Qiao, L., Du Su, F., Mao, Y.: Assessing the effect of scaling methods on retrieval of soil moisture based on MODIS images in arid regions. Toxicol. Environ. Chem. **98**(3/4), 410–418 (2016)
293. Li, B., Ti, C., Zhao, Y., Yan, X.: Estimating soil moisture with Landsat data and its application in extracting the spatial distribution of winter flooded paddies. Remote Sens. **8**(1), 38 (2016)
294. Haubrock, S.-N., Chabrillat, S., Kuhnert, M., Hostert, P., Kaufmann, H.: Surface soil moisture quantification and validation based on hyperspectral data and field measurements. J. Appl. Remote Sens. **2**(1), 23552 (2008)
295. Lausch, A., Zacharias, S., Dierke, C., Pause, M., Kühn, I., Doktor, D., Dietrich, P., Werban, U.: Analysis of vegetation and soil patterns using hyperspectral remote sensing, EMI, and gamma-ray measurements. Vadose Zone J. **12**, 15 (2013)
296. Townshend, J.R., Justice, C.O., Skole, D., Malingreau, J.P., Cihlar, J., Teillet, P., Sadowski, F., Ruttenberg, S.: The 1 km resolution global data set: Needs of the International Geosphere Biosphere Programme. Int. J. Remote Sens. **15**, 3417–3441 (1994)
297. Kustas, W.P., French, A.N., Hatfield, J.L., Jackson, T.J., Moran, M.S., Rango, A., Ritchie, J.C., Schmugge, T.J., Moran, M.S., Rango, A., Ritchie, J.C., Schmugge, T.J.: Remote sensing research in hydrometeorology. Photogramm. Eng. Remote Sens. **69**(6), 631–646 (2003)
298. Karnieli, A., Agam, N., Pinker, R.T., Anderson, M., Imhoff, M.L., Gutman, G.G., Panov, N., Goldberg, A.: Use of NDVI and land surface temperature for drought assessment: Merits and limitations. J. Clim. **23**, 618–633 (2010)
299. Müller, B., Bernhardt, M., Schulz, K.: Identification of catchment functional units by time series of thermal remote sensing images. Hydrol. Earth Syst. Sci. **18**(12), 5345–5359 (2014)
300. Krajewski, W.F., Anderson, M.C., Eichinger, W.E., Entekhabi, D., Hornbuckle, B.K., Houser, P.R., Katul, G.G., Kustas, W.P., Norman, J.M., Peters-Lidard, C., Wood, E.F.: A remote sensing observatory for hydrologic sciences: A genesis for scaling to continental hydrology. Water Resour. Res. **42**(7), 1–13 (2006)
301. Li, Z.-L., Becker, F.: Feasibility of land surface temperature and emissivity determination from AVHRR data. Remote Sens. Environ. **43**, 67–85 (1993)
302. Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A., Dech, S.: Urban footprint processor – Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission. IEEE Geosci. Remote Sens. Lett. **10**(6), 1617–1621 (2013)
303. Felbier, A., Esch, T., Heldens, W., Marconcini, M., Zeidler, J., Roth, A., Klotz, M., Wurm, M., Taubenböck, H.: The global urban footprint – Processing status and cross comparison to existing

human settlement products. In: IEEE Int. Geosci. Remote Sens. Symp. pp. 4816–4819. (2014)
304. Parcak, S., Gathings, D., Childs, C., Mumford, G., Cline, E.: Satellite evidence of archaeological site looting in Egypt: 2002–2013. Antiquity **90**, 188–205 (2016)
305. Miller, W.C.: Uses of aerial photographs in archaeological field work. Am. Antiq. **23**(1), 46–62 (1957)
306. Weber, S.A., Yool, S.R.: Detection of subsurface archaeological architecture by computer assisted airphoto interpretation. Geoarchaeology **14**(6), 481–493 (1999)
307. Netzband, M., Meinel, G., Lippold, R.: Classification of settlement structures using morphological and spectral features in fused high resolution satellite images (IRS-1C) Int. Arch. Photogramm. Remote Sens. **32**(7-4-3 W6), 160–166 (1999)
308. Ricchetti, E.: Application of optical high resolution satellite data for archaeological prospection over Hierapolis (Turkey) IEEE Trans. Geosci. Remote Sens. **62**, 3898–3901 (2004)
309. Golden, C., Murtha, T., Cook, B., Shaffer, D.S., Schroder, W., Hermitt, E.J., Firpi, O.A., Scherer, A.K.: Reanalyzing environmental LIDAR data for archaeology: Mesoamerican applications and implications. J. Archaeol. Sci. Rep. **9**, 293–308 (2016)
310. Hutson, S.R.: Adapting LIDAR data for regional variation in the Tropics: A case study from the Northern Maya Lowlands. J. Archaeol. Sci. Rep. **4**, 252–263 (2015)
311. von Schwerin, J., Richards-Rissetto, H., Remondino, F., Grazia Spera, M., Auer, M., Billen, N., Loos, L., Stelson, L., Reindel, M.: Airborne LIDAR acquisition, post-processing and accuracy-checking for a 3D WebGIS of Copan, Honduras. J. Archaeol. Sci. Rep. **5**, 85–104 (2016)
312. Parcak, S.H.: Satellite Remote Sensing for Archaeology. Routledge, New York, p. 292 (2009)
313. Parcak, S.H.: GIS, Remote Sensing, and Landscape Archaeology. Oxford Univ. Press, Oxford (2017)
314. IFRC: What is a disaster? (2019) https://www.ifrc.org/en/what-we-do/disaster-management/about-disasters/what-is-a-disaster/, Accessed 19 Feb 2019
315. United Nations: UNISDR Terminology on Disaster Risk Reduction. UNISDR, Geneva (2009) https://www.unisdr.org/files/7817_UNISDRTerminologyEnglish.pdf
316. WHO: Disasters & Emergencies – Definitions: Training Package (2002) http://apps.who.int/disasters/repo/7656.pdf
317. Benzi, E., Shurmer, I., Policella, N., Troendle, D., Lutzer, M., Kuhlmann, S., James, M.: Optical inter-satellite communication: The Alphasat and Sentinel-1A in-orbit experience. In: AIAA SpaceOps Conf. p. 13. (2016) https://doi.org/10.2514/6.2016-2389
318. Berk, A., Anderson, G.P., Bernstein, L.S., Acharya, P.K., Dothe, H., Matthew, M.W., Adler-Golden, S.M., Chetwynd Jr., J.H., Richtsmeier, S.C., Pukall, B., Allred, C.L., Jeong, L.S., Hoke, M.L.: MODTRAN4 radiative transfer modeling for atmospheric correction. Proc. SPIE (1999) https://doi.org/10.1117/12.366388
319. Berk, A., Anderson, G.P., Acharya, P.K., Chetwynd Jr., J.H., Bernstein, L.S., Dothe, H., Matthew, M.W., Adler-Golden, S.M., Richtsmeier, S.C., Pukall, B., Allred, C.L., Jeong, L.S., Hoke, M.L.: MODTRAN4 User's Manual. Air Force Research Laboratory, Hanscom AFB (1999) ftp://ftp.pmodwrc.ch/pub/Vorlesung%20K+S/MOD4_user_guide.pdf (last access: 10 February 2019)
320. Braun, D., Damm, A., Paul-Limoges, E., Revill, A., Buchmann, N., Petchey, O.L., Hein, L., Schaepman, M.E.: From instantaneous to continuous: Using imaging spectroscopy and in situ data to map two productivity-related ecosystem services. Ecol. Indic. **82**, 409–419 (2017) doi:https://doi.org/10.1016/j.ecolind.2017.06.045
321. Henderson, F.M., Lewis, A.J.: Principles and applications of imaging radar. In: Ryerson, R.A. (ed.) Manual of Remote Sensing, vol. 2, p. 896. Wiley, New York (1983)
322. Jagdhuber, T., Hajnsek, I., Papathanassiou, K.P.: An iterative generalized hybrid decomposition for soil moisture retrieval under vegetation cover using fully polarimetric SAR. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **8**(8), 3911–3922 (2015)
323. Pause, M., Schweitzer, C., Rosenthal, M., Keuck, V., Bumberger, J., Dietrich, P., Heurich, M., Jung, A., Lausch, A.: In Situ/Remote Sensing Integration to Assess Forest Health – A Review. Remote Sens. **8**, 471 (2016) doi:https://doi.org/10.3390/rs8060471
324. ISO/TS 19159-1:2014: Geographic Information – Calibration and Validation of Remote Sensing Imagery Sensors and Data. ISO, Geneva (2014)
325. ISO 19115-2:2019: Geographic Information – Metadata – Part 2: Extension for Acquisition and Processing. ISO, Geneva (2019)
326. Glossary. https://podaac.jpl.nasa.gov/Glossary

Erik Borg Erik Borg received his PhD in Natural Science from the University Potsdam, Germany, in 2007. He works at the German Aerospace Center in Neustrelitz, Germany. Since 2008 he has been lecturing on remote sensing at the University of Applied Sciences, Neubrandenburg. He works on the development of evaluation algorithms as well as in-situ measurement strategies for validating remote sensing data.

Sina C. Truckenbrodt Sina Truckenbrodt studied mathematics and physics at the University of Jena before joining the Research Groups on Earth Observation and Citizen Science at the University of Jena and DLR Jena, respectively. Currently she is a PhD candidate in Earth Observation, specializing on the exploitation of data assimilation for physically based retrieval of vegetation properties from remote sensing data acquired by optical sensors.

Angela Lausch Angela Lausch received her PhD from the University of Bonn in 2000 and her Habilitation degree in 2015 from HU Berlin (2015). Since 1993 she has been a Senior Scientist at UFZ in Leipzig and lecturer at HU Berlin. She has experience in biology, ecology, environmental health, remote sensing, and data science. She is active in numerous international networks like GEO BON.



Peter Dietrich Peter Dietrich is a geophysicist and Head of Monitoring and Exploration Technology at UFZ in Leipzig (since 2005) and also Professor of Environmental and Engineering Geophysics at the University of Tübingen (since 2012). His research focuses on the development and evaluation of measurement and monitoring methods from geophysics, hydrogeology, and geo-technology for the efficient, high-resolution investigation of the subsurface.

Karsten Schmidt Karsten Schmidt received his PhD in Physics from the University of Greifswald, Germany, in 1992. He worked at the Institute of Physics at the University of Greifswald and at the German Aerospace Center in Neustrelitz, Germany. He now works on electromagnetic and acoustic scattering on small particles and particle ensembles and on wind retrieval within ESA's wind lidar mission AEOLUS.

Surveying

11

Vladimir Golubev

Contents

Abstract

Surveying is one of the key geodetic methods, and is applied for a range of purposes: topographic maps; engineering surveying, which deals with, e.g., the construction of buildings, bridges, and dams; town and regional planning; and the development and maintenance of a property cadaster. Surveying involves the precise monitoring of deformations of natural as well as manmade structures. The variety of tools and methods used for surveying include terrestrial observations of angles and distances, the measurement of point clouds with laser scanners, photos with cameras, and combinations of these devices mounted stationary or on drones and other mobile platforms. This chapter describes the functionalities and application of the most common instruments applied in surveying as well as the integration of the resulting data into a modern GIS.

Keywords

surveying · theodolite · angle measurement · GNSS · distance measurement · leveling

The International Federation of Surveyors adopted the following definition [1]:

> A surveyor is a professional person with the academic qualifications and technical expertise to conduct one, or more, of the following activities:
>
> - to determine, measure and represent land, three-dimensional objects, point-fields and trajectories;
> - to assemble and interpret land and geographically related information,
> - to use that information for the planning and efficient administration of the land, the sea and any structures thereon; and,
> - to conduct research into the above practices and to develop them.

The goal of surveying is the determination of the coordinates of points in 2-D or 3-D using terrestrial or satellite methods.

Methods of surveying can be subdivided into the measurement of angles, distances, single points, and point clouds.

A *theodolite* or a *total station* performs angle measurements. A theodolite can measure angles only, while a total station can measure angles and distances. In both cases, horizontal and vertical angles are measured by separate components of the instrument. An angle measurement requires a direction to a reference point followed by a direction to one or directions to many new points. The measurement error of the most accurate instruments used today is a half of a second ($0.5''$) or $1/2\,592\,000$ of a full circle (about 0.15 mgon). This corresponds to about 2.5 mm at a distance of 1 km. The mea-

V. Golubev (✉)
Faculty of Territory Development, Moscow State University of Geodesy and Cartography
Moscow, Russian Federation
e-mail: golubev@miigaik.ru

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_11

surements are stored in the instrument's internal computer. This internal computer often provides a full set of programs for geodetic applications.

Except for the introduction of electronic angle detection, registration, and storage, theodolites have not changed much for more than century, although their accuracy has improved over the years.

Distance measurements are mostly performed with electronic distance meters. There are two methods – Pulse and phase. In pulse method the distance is calculated using the speed of light and the travel time of light, which are both known. In phase method, an instrument sends a carrier wave with a superimposed reference wave (that has a wavelength of, e.g., 10 m) to an object and then measures the difference of phases between sent and reflected signal. The number of full wavelengths plus the fragment of the last wavelength defines the distance. The number of full wavelengths is calculated using additional information. For example, information which get by using an another wavelength. The maximum distance that can be measured using this basic method is several kilometers, with an accuracy of, at best, 0.2 mm + 0.2 ppm (achieved using a prism as the reflector). The wavelength of the carrier wave—produced by a laser—is short, i.e., in the range of about 1 μm (10^{-6} m) or shorter.

Electronic distance meters are either devices built into a total station or handheld instruments for performing measurements at buildings.

In the past, several methods were used for distance measurements. Examples include chains, a surveyor's rod, measuring tapes, and combinations of angle measurements and rods. The measurement tape is still used for special applications. All the other methods have been consigned to history.

The leveling method is used to measure geodetic heights. Here, the height difference between two points is determined as follows. First, a horizontal line of sight is realized by the telescope of the leveling instrument. Second, determination the height difference between each of the two points using horizontal line of level.

Single points can be directly measured with a global navigation satellite system (GNSS). The basic measurements are then the four distances to four satellites. Any distance corresponds to a certain number of full wavelengths plus a fragment of a wave. The number of full wavelengths is determined with a time measurement such as that performed by electronic distance meters. The fragment is determined by an interferometric comparison with a full reference wave. The wavelengths used in a GNSS are about 20 cm. This method is used for navigation purposes and has an accuracy of at least several meters.

A point on the Earth is measured as the intersection of three spheres centered on the known positions of three of the satellites. The fourth satellite is necessary to determine the error in the receiver's clock.

The geometric model of a spatial arc section assumes that the waves follow straight lines. However, refraction in the atmosphere causes slightly curved lines. This unwanted influence can be estimated by performing two parallel measurements with two slightly different frequencies (and thus wavelengths) and using an atmospheric model, and can hence be eliminated in a way that makes the GNSS applicable for most applications. The error is no more than 5 m.

For more precise measurements, a second receiver must be placed at a known point in the neighborhood of the measurement. Then, the remaining atmospheric error is computed as the difference between the given point coordinates (target) and the measured point coordinates (actual value). This difference is applied with a negative sign to the new point. This method is called differential GNSS (DGNSS). To get even higher accuracy, the phase difference of the carrier wave is measured, which leads to an accuracy of a few centimeters or better if the new point is closer than about 30 km to the known point. For measurement, waves of the same length, at the same time, are generated on the satellite and on the receiver. The satellite sends a signal to the receiver. Since the signal travels the distance from the satellite to the receiver, the phase of the received signal will be different from that generated. This phase difference is measured.

Laser scanners and the dense-stereo-matching method of photogrammetry generate point clouds. A point cloud is a large set of 3-D points that are placed on the surface of an object. Every point has the coordinates x, y, z and, eventually, color information such as R, G, and B, or at least intensity. Several point clouds are referenced together using coordinate transformation with sufficient identical control points.

Point clouds can be fairly quickly generated and often deliver a good visualization of the object. However, most other applications require information that is more concise. Therefore, there are algorithms that automatically extract edges, planes, and other geometric elements from a point cloud. Currently, however, these algorithms give rather poor results.

11.1 Surveying Instruments

The measurements carried out by geodetic instruments allow the positions of points and other spatial objects to be defined, derived products such as areas to be calculated, and investigations of engineering constructions (regarding their deformation, movement, and other parameters) to be performed.

Usually, geodetic measurements encompass angles, distances, levels, determinations of the positions of points, and parameters of the gravimetric field of the Earth.

The various measurements are described in the following sections.

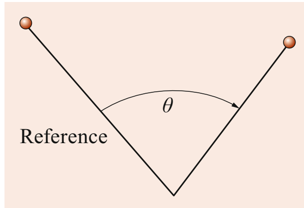


Fig. 11.1 Basic requirements of an angle

11.2 Angle Meters and Measurement

Many geodetic computations depend on the measurement of angles and directions.

In geodesy, angles are classified as horizontal or vertical angles depending on the plane in which they are measured.

Three basic requirements determine an angle θ. They are the reference or starting line, the turning direction, and the angular distance (Fig. 11.1).

Azimuths are horizontal angles observed clockwise from any reference meridian, which is usually north in plane surveying. Depending on the reference meridian, azimuths may be geodetic, grid, magnetic, or others. A geodetic azimuth refers to the direction to the North Pole. A grid azimuth refers to the northern end of a line parallel with the central meridian. A magnetic azimuth refers to the direction to magnetic north (Fig. 11.2).

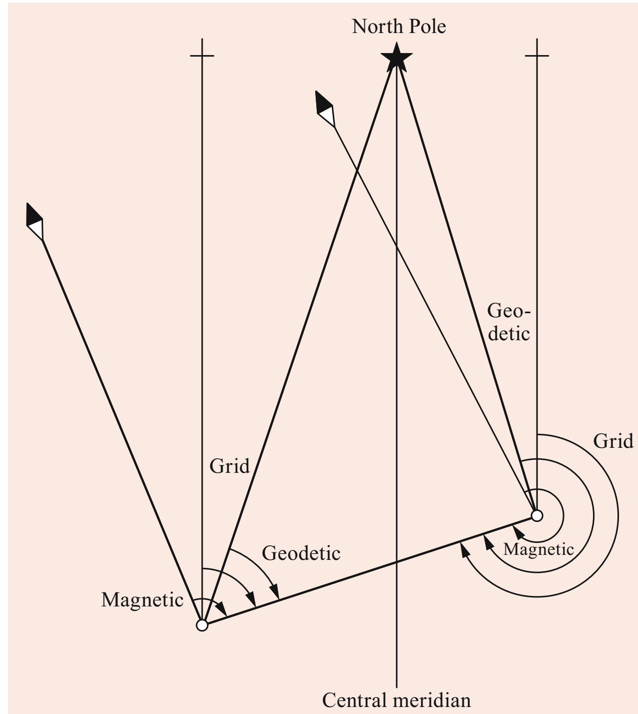


Fig. 11.2 Geodetic, grid, and magnetic azimuths

11.2.1 Theodolites

In geodesy, theodolites and total stations are used for the measurement of horizontal and vertical angles.

The theodolite is the geodetic instrument intended for the measurement of angles. It can be either purely optic or opto-electronic. In modern geodesy, electronic devices record data in a digital form, allowing smooth dataflow to databases and subsequent processes.

Nowadays, however, theodolite usage is becoming increasingly rare due to the popularity of modern total stations.

Depending on their accuracy, theodolites are subdivided into high-precision (RMSE 0.5–1″, where RMSE is the root-mean-square error), exact (RMSE 2–5″), and technical (RMSE 10–60″).

For example, theodolites applied in construction works belong to the last group. A schematic of a theodolite is given in Fig. 11.3.

A theodolite has circular glass limbs divided into 360° or 400 gon steps for the measurement of horizontal and vertical angles. During such measurements, the limb remains in the fixed position. Around the horizontal limb, there is a rotating circular glass plate—the alidade, which is rigidly connected to the column of the theodolite. A telescope can rotate around the axis HH′. The axis of sight of the telescope is designated VV′.

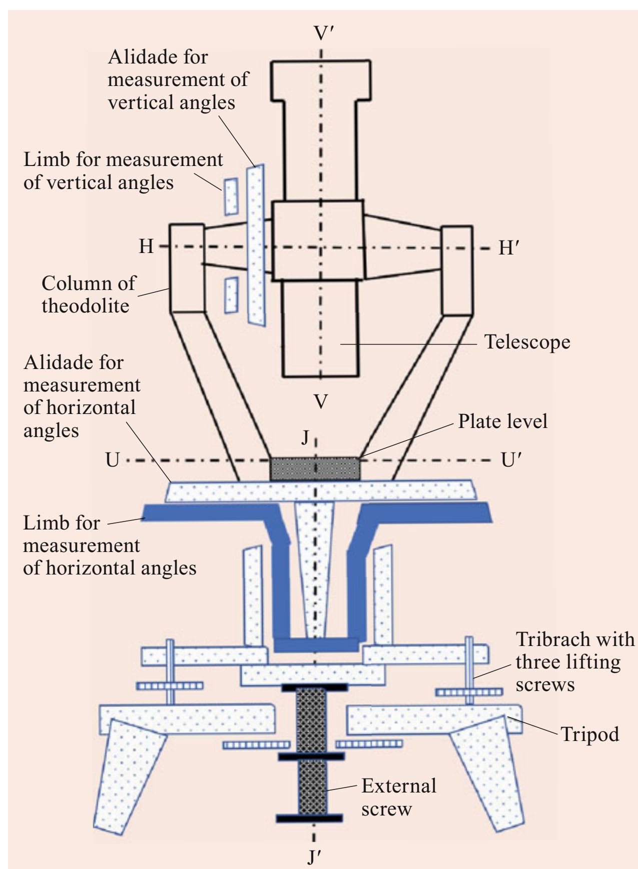


Fig. 11.3 Schematic of the components of a theodolite

11

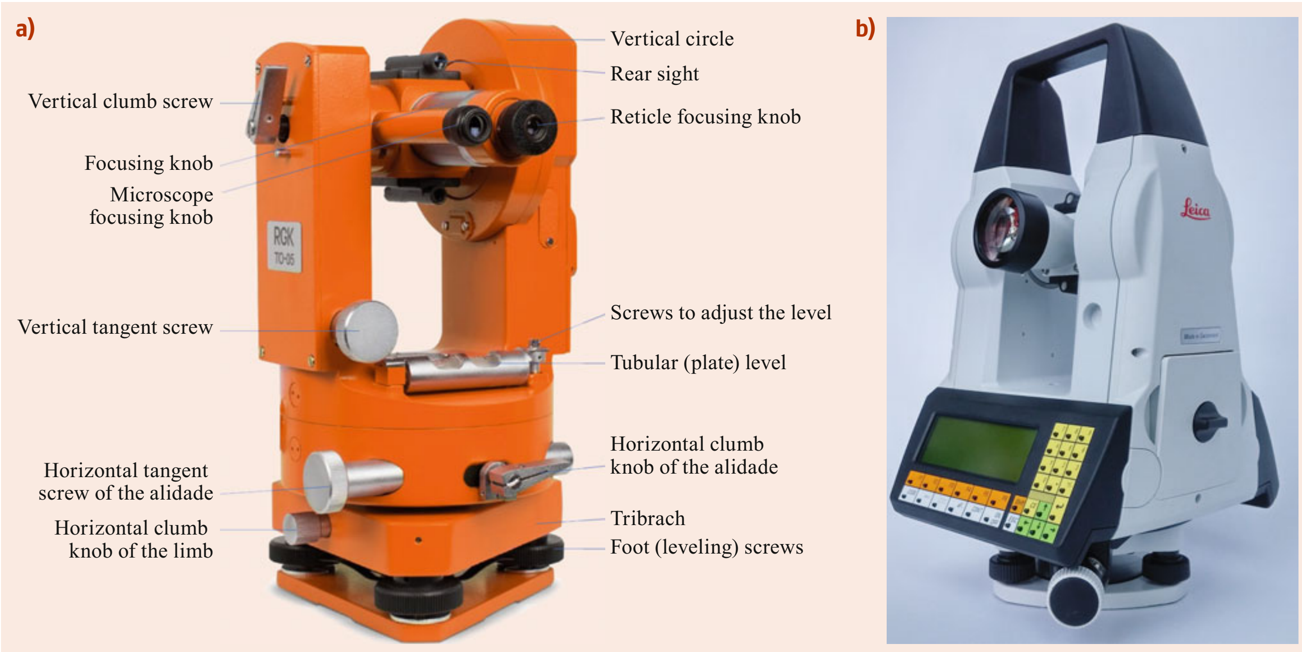


Fig. 11.4 **a** Photograph of an older theodolite with many visible components. **b** The Leica TM 5000, a modern computer-controlled theodolite (courtesy of Leica)

The axis JJ′ is the vertical axis of rotation of the theodolite.

The theodolite is set to the horizontal position by means of three lifting screws and a cylindrical level. The vertical circle is rigidly connected to the axis of rotation of the telescope. It can be located on the right or left side of the telescope. The first situation is called *vertical circle left* (face left—FL) and the second *vertical circle right* (face right—FR). The vertical circle also consists of a limb (the fixed part) and an alidade (the rotating part). The theodolite is fastened to a tripod by means of an external screw.

Photographs of theodolites are given in Figs. 11.4a and11.4b. The rotating parts of a theodolite are supplied with clamp screws to fix those parts in position, and tangent screws for the exact orientation of the device in the set direction.

11.2.2 Measurement of Horizontal Angles with a Theodolite

The theodolite must be exactly centered over station mark A before beginning the observations (Fig. 11.5a). The vertical axis must be set in the plumb position. In this case, the limb and alidade will be in the horizontal plane. Thus, the vertical axis of rotation will coincide with the plumb line to the survey point of the instrument. The angle will be measured in the horizontal plane. The method of receptions is applied to measure one angle. If it is necessary to measure several directions from point A, the method of *circle receptions* is applied. In the latter case, observations are done in circular order. The first observation is directed to the rear point B. Then the clamp screws of the limb and the alidade are fixed. Exact targeting of the telescope is performed with the tangent screw. The observation of point B at the limb may be called b. After this observation, the alidade is unfastened and the telescope pointed at point C in order to observe c. Consequently, angle β between B and C is the difference $(b - c)$. This sequence characterizes the measurement of an angle by one *semi-reception*. A *full reception* consists of two semi-receptions. First, the angle is measured in the vertical circle left position (first semi-reception; Fig. 11.5b). Then the telescope is turned 180°. The vertical circle appears to the right of the telescope (Fig. 11.5c). Next, the measurement of the angle is repeated. This is the second semi-reception. The difference between the measured angles for the two semi-receptions must be smaller than twice the mean square deviation (for the theodolite). For example, angle measurements performed with a $2''$ theodolite must result in a difference that is smaller than $4''$. The final value of the angle is the arithmetic average.

If it is necessary to measure angles between several directions at the station, the *closing of horizon* method is applied (Fig. 11.6). Here, the position of the limb is fixed and the telescope is directed clockwise (the clockwise case) such that it is pointed at points 1, 2, 3, 4, and 5 and then again at 1, performing a direction measurement at each point. Then the position is changed to circle right and the measurements are repeated counterclockwise. Finally, a full reception of measurements of all angles at the station is completed.

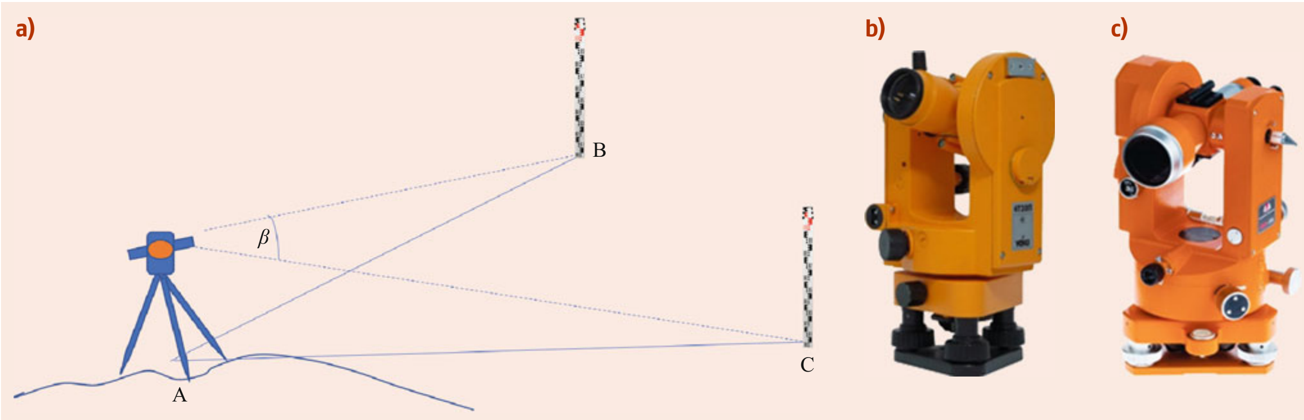
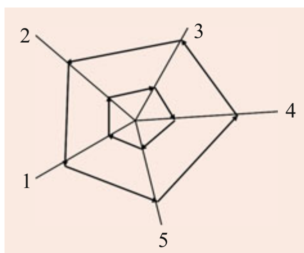


Fig. 11.5 **a** Measurement of horizontal angles. **b** Vertical circle left. **c** Vertical circle right

Fig. 11.6 Closing of horizon



11.2.3 Classifications of Theodolites

Usually, theodolites are categorized by their working principle and accuracy. The working principle of a theodolite may be optical or electronic.

The main advantage of an optical theodolite is full autonomy and its lack of dependence on power sources. Moreover, in view of the absence of electronic components, it is capable of working at extremely low temperatures, down to −30 to −40 °C!

The representation of measured values in digital form is the cornerstone of the work of an electronic theodolite. The internal memory avoids the need for multipage field magazines and allows measurements to be recorded directly into the memory of the device, increasing the speed of work and reducing the number of mistakes.

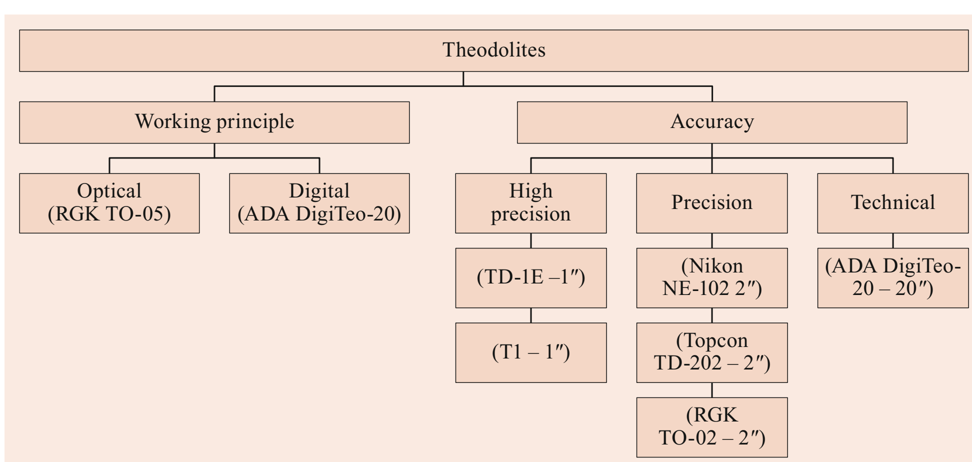
Figure 11.7 shows the classification of theodolites by working principle and accuracy. It also includes examples of specific theodolites belonging to the shown groups.

11.3 Distance Measurement

Surveying offers many methods for performing linear measurements. They include pacing (counting the number of steps), odometer readings (counting the number of revolutions of a wheel and multiplying by its circumference), taping (using a measurement tape), electronic distance measurement (EDM), and distance computation based on point measurements made by global navigation satellite systems (GNSSs).

The accuracy of pacing is sufficient for many applications, e.g., in forestry or to detect blunders. An odometer attached

Fig. 11.7 Classifications of theodolites (examples are shown in brackets)



to a vehicle supports route-location works. Taping is still in use for smaller tasks. However, EDM and GNSSs are by far the most widely applied methods because of their accuracy and efficiency. In addition, GNSSs allow the measurement of long distances.

11.3.1 Principle of Electronic Distance Measurement

The principle of electronic distance measurement by phase method is depicted in Fig. 11.8 [2]. The EDM device (sender and receiver) is centered over point A and a reflector is centered over point B. A reference frequency with a precisely controlled wavelength is superimposed on a carrier wave. The reflector at B returns the signal to the receiver. Thus, the travel path is twice the slope distance between A and B.

The case where there is an exact number of wavelengths within the distance $\overline{AB}$ is obviously a theoretical one. Usually, a fractional part p of a wavelength, as shown in Fig. 11.8, is needed to complete the distance. In Fig. 11.9, the distance L between the EDM instrument and the reflector would be expressed as

$$L = \frac{n\lambda + \mathrm{p}}{2}, \tag{11.1}$$

where λ is the wavelength, n the number of full wavelengths within the travel path, and p the length of the fractional part.

The fractional part p is determined by the EDM instrument from a measurement of the phase shift of the return signal.

Example: $\lambda = 20\,\mathrm{m}$. The phase shift of the return signal is 123.4°, and the number of full wavelengths $n = 11$ (Fig. 11.9).

$$p = \frac{123.4°}{360°} 20.000\,\mathrm{m} = 6.856\,\mathrm{m} \tag{11.2}$$

$$L = \frac{11 \cdot 20.000\,\mathrm{m} + 6.856\,\mathrm{m}}{2} = 113.428\,\mathrm{m}. \tag{11.3}$$

Not all EDM instruments require reflectors for distance measurement. Those instruments use time-pulsed infrared laser signals, and can observe distances of up to 200 m in the reflectorless mode (Fig. 11.10).

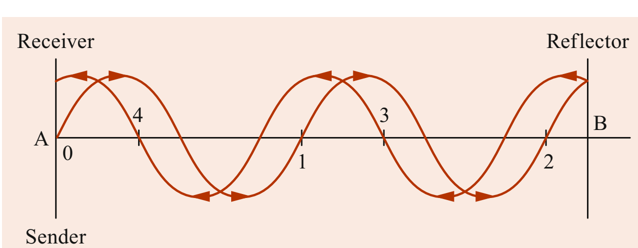


Fig. 11.8 Principle of electronic distance measurement

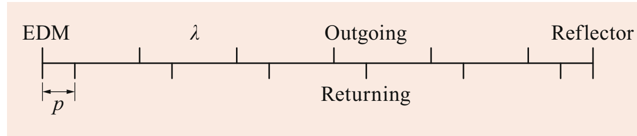


Fig. 11.9 Range between the EDM and the reflector, showing the fractional part p of a wavelength

11.3.2 Total Stations

The difference between total stations (tacheometers) and theodolites is the built-in distance meter of a total station. This allows the measurement of not only angles but also distances. A total station consists of an angle-measuring component such as a theodolite, an EDM instrument, and a computer that serves for data registration and further processing.

A total station can upload any application program. This allows the automation of measurement processes and the execution of geodetic tasks within the instrument itself, thus expanding the operational possibilities and field of applications of the total station.

The automated registration capability allows the creation of a fully automated geodetic workflow: registration of information–conversion–computing–plotting. Thus, the outcome is an end product such as a digital topographical plan. At the same time, errors introduced by the observer, the operator, and the cartographer are minimized at each stage of work compared to the traditional approach.

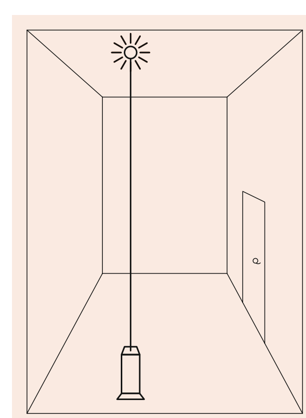
A small number of companies offer total stations. The best known are Leica, Trimble, Topcon, and Nikon.

The general scheme for and an example of a total station are provided in Figs. 11.11a [3] and11.11b, respectively.

The classification of total stations is presented in Fig. 11.12.

Modern digital electronic tacheometers integrate an electronic theodolite, a high-precision distance meter, and a field computer. The angle reading unit is mostly based on barcodes (Fig. 11.13). The limbs of an electronic theodolite

Fig. 11.10 Principle of a reflectorless measurement



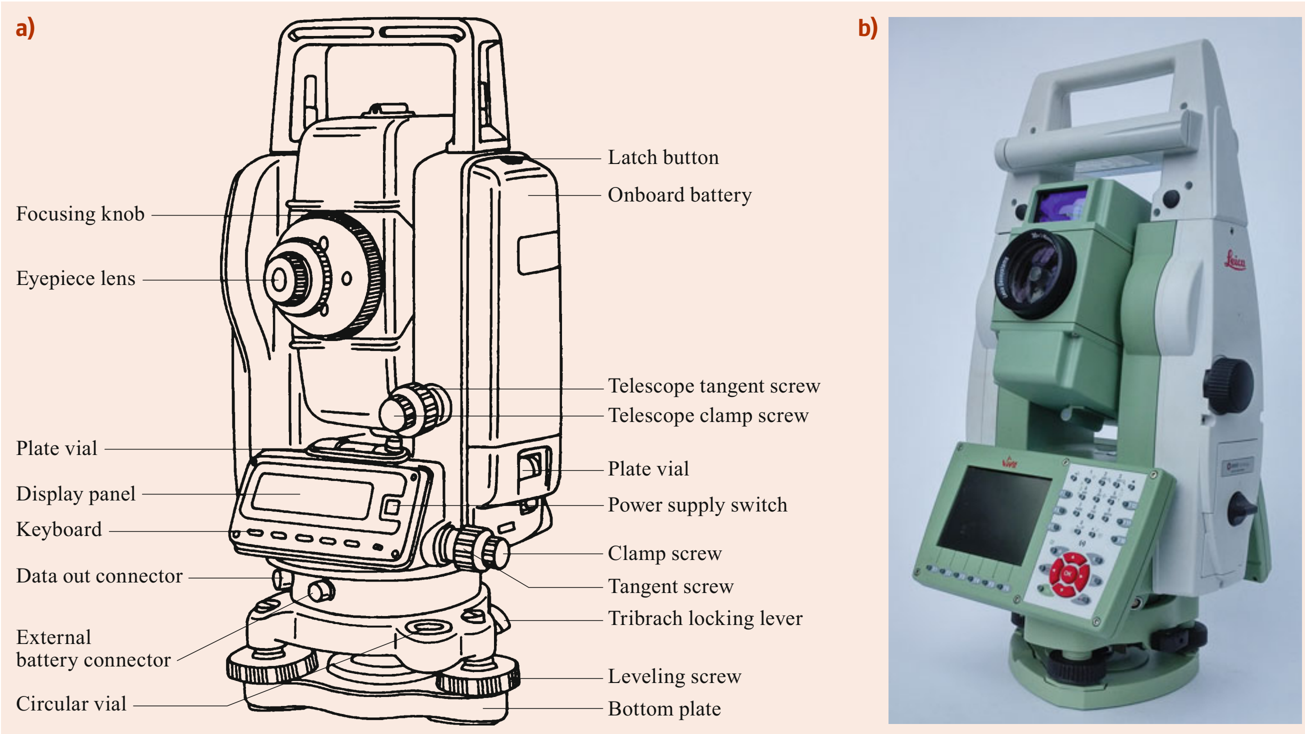
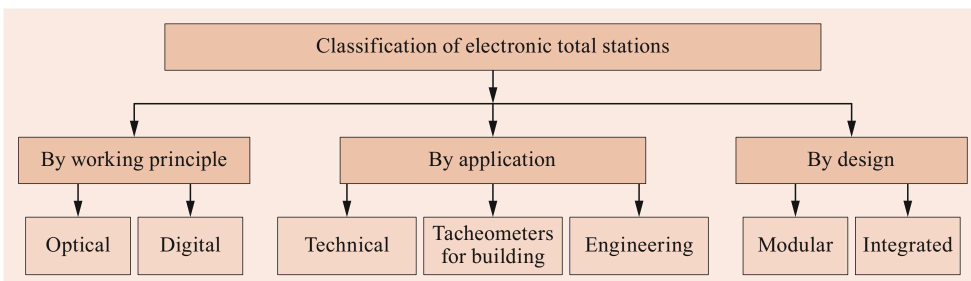


Fig. 11.11 **a** General scheme of a total station. **b** Example of a modern total station: the Leica TS 15 (courtesy of Leica)

Fig. 11.12 Classification of total stations



incremental measuring system consist of an opaque coating divided into an equally spaced sequence of dark and transparent strips. Usually, the angular interval between these strips is 1–2′. The limb also has a barcode as a reference mark—the zero reading of the limb. A barcode explicitly identifies the zero position. The readout of an angle measurement is computed on the basis of the picture of the fixed alidade (strips), as recorded by a photodiode. A light source can be found below the location at which recording occurs (Fig. 11.13) [5].

The incremental limb scale is a regular sequence of equal dark and transparent stripes. There is a light source from which light passes through the limb and fixed mask and enters the five-photodiode matrix. The mask has five transparent stencils. One of them has a bar of barcode identical to that which has a Zero barcode on the limb. When we rotate the limb, the limb strips completely coincide with the mask strip. This is a complete superposition. And the zero photodiode generates

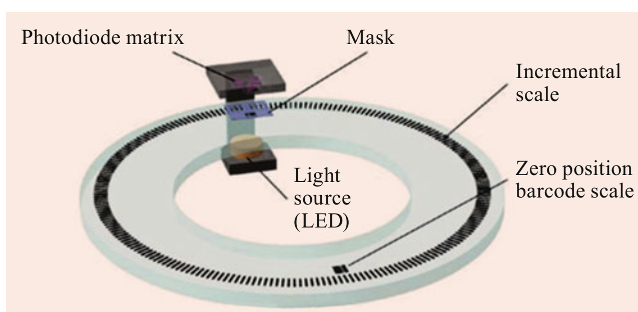


Fig. 11.13 Incremental measuring system (graduated circles)

a short pulse. This is the moment when the zero reading is fixed. Further, when the limb rotates, it is fixed number of limb strips, which corresponds to the measured angle.

There is a barcode scale for the limb circumference in the positional type system. Such a limb has an endless bar of barcode spaced circumferentially. It determines the count. In this case, there is no zero value. Anywhere in the limb there is a unique barcode that determines the limb count. Thus, it is not necessary to have a zero barcode. Therefore, this reading method is called the absolute method. Barcodes correspond to angular readings on the limb. As mentioned above, this device consists of a light source and a CCD array (dedicated analog integrated circuit). The absolute angle sensor consists of an LED and a CCD line, where bar code band images corresponding to the measured angle are projected and read.

In order to eliminate the systematic errors from a total station, either a special measurement procedure can be performed or built-in software can be applied to eliminate those errors by computing corrections. For example, the eccentricity error can be eliminated by transiting the total station and then performing a second readout. In this case, the software of the total station processes the two readouts and averages the received values. If a unilateral reader is used, the amendment depends on the internal software of the total station. Because of the multiplicity of total station types, the technical characteristics of a specific device need to be retrieved from its documentation.

Distance meters of total stations can be adapted to different reflectance technologies: a reflector prism, a reflecting plate or film, or none of these (only the surface of the object of interest). Distance meters apply the impulse or the phase method for distance measurement. The impulse method uses the time between the ignition of the laser pulse and the detection of its reflection after it has traveled twice the distance. The phase method uses the difference between the phase positions of the radiated and received waves.

Total stations can be subdivided into the following groups based on their application:

- *Technical total stations* contain a basic set of functions. They are intended to complete simple and routine tasks. Such total stations are only equipped with a reflector. The workmen required are the operator of the total station and the rodman holding the reflector.
- *Building total stations*. Such instruments are used for the geodetic maintenance of construction works, and usually consist of electronic distance meters without a reflector.
- *Engineering total stations*. These have extended functionality so that they can be applied to different types of surveying and complex geodetic work. This functionality may include a camera, modules for the creation of three-dimensional models, a touch screen, a powerful processor, USB ports, and access to Wi-Fi and Bluetooth. Only a small proportion of all total stations are engineering total stations.

Total stations can also be subdivided based on their design. Modular devices consist of several independent elements; for example, a theodolite and nozzle for distance measurements. Integrated total stations include several mechanisms in one case, and can be motorized and automated.

The majority of the total stations used today have a laser plumb for centering the device. The typical power supply of a total station is a lithium-ion rechargeable battery. The result of a measurement is shown on a liquid crystal screen and can be registered in the internal memory, which typically has a capacity of 10–20 thousand measurements. Aside from the measurements, the most important parameters (e.g., the type of distance-measurement mode, the battery charging level, etc.) can be constantly displayed on the screen. Nowadays, there is a large class of robotic total stations that can be operated by a single person. Such devices combine a servo-driver with special systems that are capable of tracing an object, identifying it, and obtaining the required data. Robotic total stations allow remote observations.

The accuracy of angular measurements by modern total stations can reach half an angular second ($0°00'00.5''$). This is equivalent to an accuracy of 2.4 mm per kilometer.

11.4 Leveling

Leveling is a general term that denotes the measurement of elevations and differences in elevation. The key terms involved are defined below [4]:

- *Level line*: a line of constant or uniform height relative to mean sea level, and therefore a curved line that is concentric with the mean surface of the Earth. More formally, a level line is a line on a level surface and is normal (at right angles) to the direction of gravity at all points along its length.
- *Horizontal line*: a line through a point tangential to the level line passing through the same point and normal to the direction of gravity at that point.
- *Vertical datum*: a level surface or line from which heights are measured, or to which heights may be referred.
- *Elevation* (height): the vertical distance of a point above or below a datum surface.
- *Mean sea level* (MSL): the mean level of the sea as determined at a selected location from observations made over a period of time, using a datum surface for leveling work.
- *Bench marks*: fixed points with heights relative to a datum surface that have been determined using a surveyor's level.
- *Leveling*: the name given to the method of determining heights, or rather differences in height, using the instrument known as the surveyor's level.

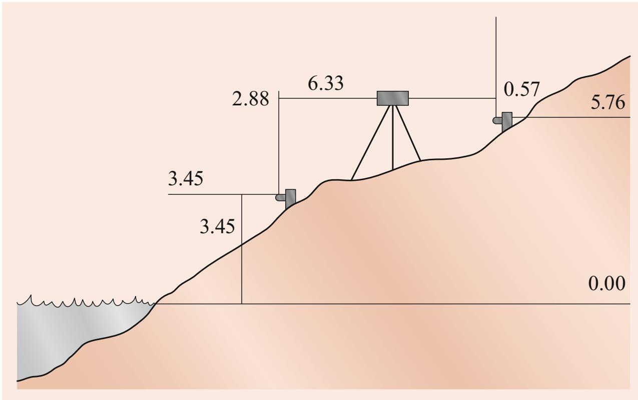


Fig. 11.14 Principle of differential leveling

11.4.1 Leveling Methods

Elevation differences may be determined by taping or electronic vertical distance measurement, differential leveling, barometric leveling, or by trigonometry.

Taping or *vertical electronic distance measurement* are applied in construction works and mining. Examples are height measurements for rooms and depth measurements for shafts.

Differential leveling is the most common leveling method. The crosshair inside the leveling instrument defines a line of sight that is exactly aligned with a horizontal line by a manually adjusted level vial or an automatic compensator. The instrument is placed approximately at the midpoint between two vertical rods. The difference between the two readings is the difference in elevation between the two rod positions. Figure 11.14 gives an example where the difference in elevation is $2.88 - 0.57 = 2.31$ m.

Barometric leveling is the use of a barometer to determine an elevation difference.

A barometer measures the absolute air pressure, which depends on the weather conditions, the time of day and year, and the elevation. An increase in elevation of about 8 m causes a reduction in air pressure of about 1 hPa.

In stable weather conditions, and by using several barometers, the measurement of elevations with an accuracy of less than a meter is possible [2].

Trigonometric leveling is the measurement of the vertical angle ν (Fig. 11.15).

It is possible to calculate the difference in height h between A and B using the formula $h = d \cdot \tan(\nu) + i - l$, where d is the horizontal distance, ν is the measured angle, i is the height of the theodolite, and l is the observation height.

The measurement of the vertical angle ν is carried out relative to the horizontal plane.

Theoretically, when the telescope is in a horizontal position, the readout at the vertical circle is expected to be zero. Usually, the level of the vertical circle that is rigidly

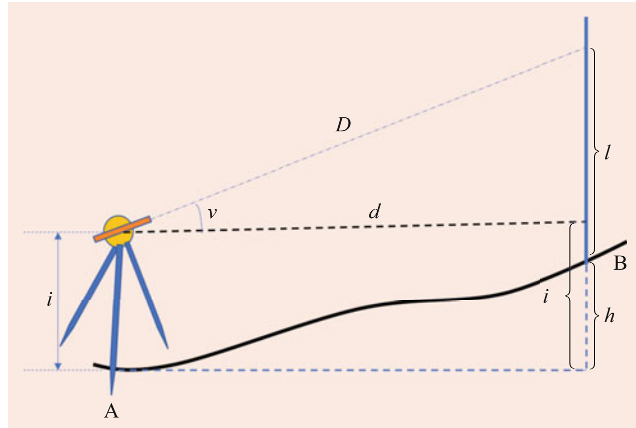


Fig. 11.15 Height measurement using trigonometry

connected with a limb is used for this purpose. Thus, while the position of the bubble of the level is centered, and the telescope is set to its horizontal position, the readout at the vertical circle has to be zero. However, because the settings of the level and the limb are not free from errors, the readout usually differs from zero. This counting is called index error (IR). Thus, the vertical angle is measured in relation to a tilted reference. The slope of the tilted reference is not known beforehand, but it can be determined during the measurement of a full reception. This means that the target must first be observed with the vertical circle left (*face left*, FL) and then with the vertical circle right (*face right*, FR), resulting in measurements to points A and B (average value, Fig. 11.16).

The tilt angle is calculated as the difference $\nu = \text{FL} - \text{IR}$ or $\nu = \text{IR} - \text{FR}$, where

$$\text{IR} = \frac{\text{FL} + \text{FR}}{2}. \tag{11.4}$$

If IR differs considerably from zero, the absolute value can be reduced by adjusting the screws of the level.

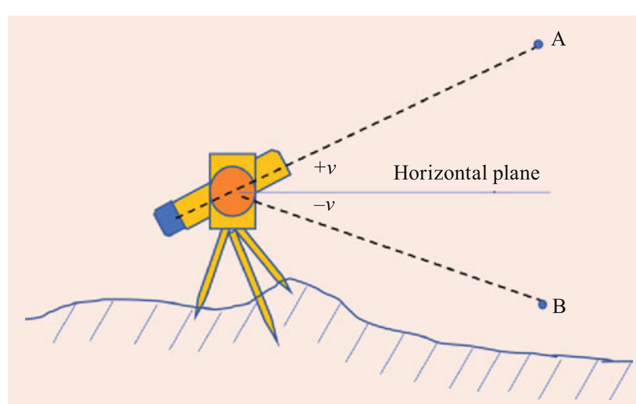


Fig. 11.16 Measurement of vertical angles

11

Fig. 11.17 Tilting level: the Wild N3 (courtesy of Trimble)

11.4.2 Leveling Equipment

The instruments used for differential leveling can be grouped into three categories: tilting levels, automatic levels, and digital levels. They all have two components in common: a telescope and a system for creating a horizontal line of sight.

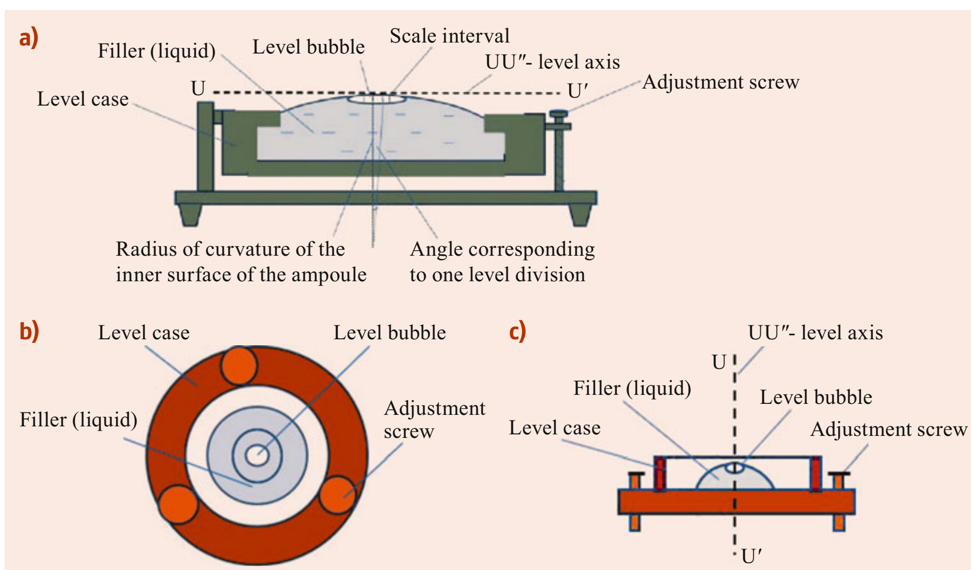
A *tilting level* is equipped with a level vial that is manually adjusted for every reading. It takes longer to use a tilting level than the other types of leveling instruments, but it is still often employed for its higher precision (Fig. 11.17).

Level vials are used to orient many surveying instruments with respect to the direction of the gravity vector. Two types are in use: the tube vial and the circular vial. The upper surface inside a vial is slightly curved. The bubble of the vial always moves to the highest point and thus indicates the tilt of the vial. A high-precision vial allows angles to be measured to an accuracy of 20″ (Fig. 11.18).

Automatic levels are equipped with a compensator to establish a horizontal line of sight without the need for manual interaction. Thus, they are much faster to use.

The operating system of one type of automatic compensator is shown in Fig. 11.19a. The system consists of two fixed prisms and a third prism suspended from wires to create a pendulum. Thus, a horizontal line of sight is achieved, even though the telescope itself may be slightly tilted away from horizontal (Fig. 11.19b; [2]).

Fig. 11.18 **a** Tube vial. **b**, **c** Circular vial. **b** Top view, **c** side view



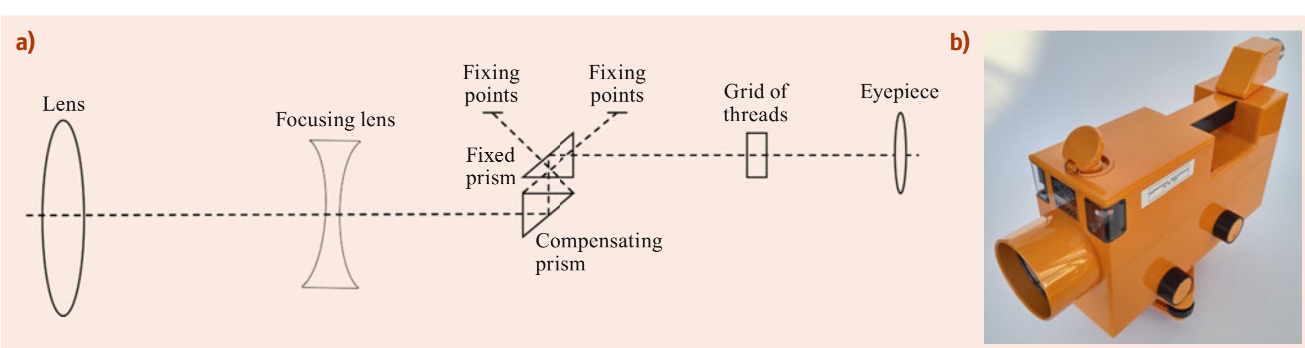


Fig. 11.19 **a** Scheme for an optical compensator. **b** Automatic level: the NI002 (courtesy of Trimble)

Fig. 11.20 **a** Digital level (DiNi 12T from Zeiss). **b** Barcoded rod (courtesy of Trimble)

Digital levels have all of the components included in automatic levels plus an image-processing component that allows automatic reading of barcoded rods. Thus, uninterrupted dataflow from the reading to a digital data store can be established (Fig. 11.20).

11.5 Global Navigation Satellite Systems

The following global navigation satellite systems (GNSSs) are currently in operation:

- GPS—NAVSTAR (Global Positioning System—Navigational Satellite Time and Ranging), or simply GPS: the American navigation system that is presently the most widely used GNSS in the world.
- GLONASS (ГЛОНАСС; Global Navigation System): the Russian navigation system.
- Galileo: a European system that partially earned in 2016. Now the Galileo constellation consists of 24 active satellites.

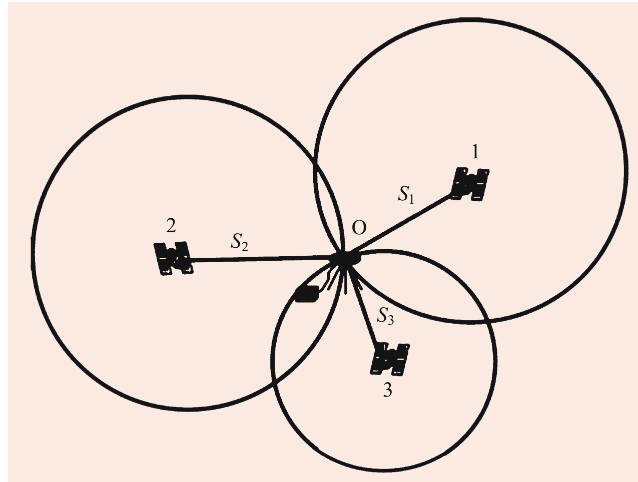


Fig. 11.21 Geometrical interpretation of a spatial linear intersection (2-D view of three satellites)

- DORIS: the French system.
- BeiDou: the Chinese system.

11.5.1 Principles of a GNSS

A GNSS is a system that is used for the positioning of objects on land, on water, and in the air, as well as some other tasks. The basic elements of a satellite system for navigation are as follows:

- An orbital group of satellites radiating special radio signals.
- An Earth-based control system (a land segment) that monitors the current positions of the satellites and transfers corrections to the satellites in orbit. This segment controls the GNSS satellites, traces them, and provides precise information on their orbits and the time.
- Equipment (a satellite receiver) that enables the user to make use of the navigation satellite system.

By means of these measurements, it is necessary to determine the location (coordinates) of the point on the ground where the user receiver is located. The positions of the satellites are known (Fig. 11.21). If we measure the distance from only one satellite to a receiver located on the ground, we can say that the receiver is located somewhere on a sphere whose radius is equal to the measured distance. At the same time, the satellite is located in the center of such a sphere. Similarly, if the distance from the second satellite to the receiver is measured, a second sphere with a radius equal to the second measured distance is obtained. And the second satellite will be at the center of the second sphere. The measurement from the third satellite will result in us getting a third sphere. The three spheres will intersect at one point. So we get the receiver's location on the ground. Analytically, the coordinates of the receiver points on the ground are determined from the following equations

$$S_1 = \sqrt{(X_1 - X_0)^2 + (Y_1 - Y_0)^2 + (Z_1 - Z_0)^2}$$
$$S_2 = \sqrt{(X_2 - X_0)^2 + (Y_2 - Y_0)^2 + (Z_2 - Z_0)^2}$$
$$S_3 = \sqrt{(X_3 - X_0)^2 + (Y_3 - Y_0)^2 + (Z_3 - Z_0)^2}, \quad (11.5)$$

where S_n are the geometric ranges from the three satellites to the receiver at station O. (X_n, Y_n, Z_n) are the geocentric coordinates of satellite n at the time of signal transmission, and (X_0, Y_0, Z_0) are the geocentric coordinates of the receiver at transmission time.

From a geometrical point of view, only three satellites are required to solve the equations. However, there is an additional unknown due to the imprecisely known time difference between the satellites' clocks and the clock of the receiver: the so-called *bias*. This bias is the same for all three ranges since the same receiver is observing each range. The use of a fourth satellite range allows the receiver's clock bias to be mathematically determined. Therefore, one point measurement requires simultaneous observations by at least four satellites.

A table of the positions of all the satellites is called an *almanac*. Knowledge of its contents is a prerequisite for the measurement of any satellite receiver. Each satellite transfers all of its almanac information as a part of its signal. Thus, it is possible to calculate the position of a point in space on the basis of the information provided by the almanac. Table 11.1 characterizes the systems GLONASS, GPS, and Galileo.

Synchronization is carried out using atomic clocks, which are available both on the satellite and on the receiver. Two methods are used for the measurement: the code method and the phase method.

In the code method, a signal is generated on the carrier frequency. Then a modulated code signal is imposed on this carrier frequency.

By means of synchronized clocks, a similar code signal is generated on the satellite and the receiver at the same time (Fig. 11.22). The receiver, which is located at the earth, accepts this satellite signal with some delay.

The period $\Delta\tau$ is the time taken for the signal to travel the distance from the satellite to the receiver. Here, T_s and T_r are the times at which a code signal is generated on the satellite and in the receiver on the Earth, respectively. The distance is then defined using the propagation speed of the radio signal, C, via

$$S = C\Delta\tau.$$

As the speed of a radio wave is about 300 000 km/s, the time must be measured to an accuracy of 10^{-10} to 10^{-12} s to ensure the geodetic accuracy of the measurement.

The most challenging task is to determine the difference between the measurements of the recordings at the receiver.

In the phase method, the phase difference between the carrier frequency signal (from the satellite) and the signal generated in the receiver (Fig. 11.23) is calculated.

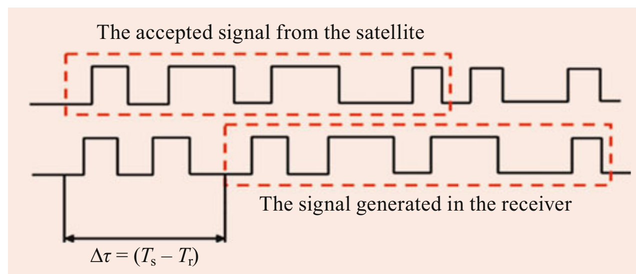


Fig. 11.22 Definition of the time of propagation of a code signal from the satellite to the receiver

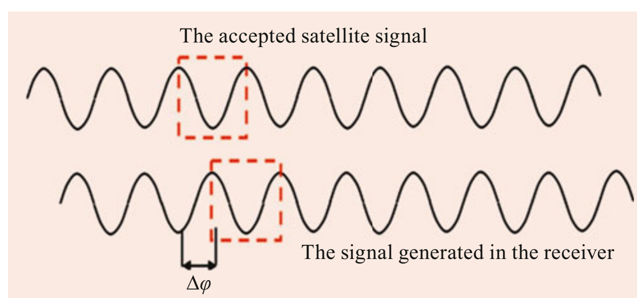


Fig. 11.23 Phase method for the determination of distance

Table 11.1 Technical characteristics of the satellite systems GLONASS, GPS, and Galileo

Characteristic	GLONASS	GPS	Galileo
Number of satellites in the full orbital group	24(3)	24(3)	27(3)
Number of orbital planes	3	6	3
Number of satellites in each plane	8	4	9
Orbital inclination	64.8°	55°	56°
Height of orbit [km]	19 130	20 200	23 222
Orbital period [hh:mm:ss]	11:15:44	11:58:02	14:04:41
Coordinate system	ПЗ-90 (PZ-90; Earth Parameters 90)	WGS-84	Galileo Terrestrial Reference Frame (GTRF)

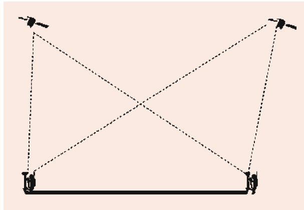


Fig. 11.24 Scheme for relative measurements from two points

The GPS system operates at two main frequencies, L1 and L2:

$$f_{L1} = 1575.42\,\text{MHz}; \quad \text{wavelength } \lambda_{L1} = 19\,\text{cm}$$
$$f_{L2} = 1227.60\,\text{MHz}; \quad \text{wavelength } \lambda_{L2} = 24\,\text{cm}$$

In the GLONASS system, each satellite has a different pair of main frequencies. They differ for L1 at a magnitude of $k \times 0.5625$ MHz with an initial frequency of 1602 MHz, and for L2, at a magnitude of $k \times 0.4375$ MHz, with an initial frequency of 1246 MHz, where k is the satellite number.

With the phase measurement method, accuracy depends on wavelengths. The shorter the wavelength (the greater the signal frequency), the more accurate the measurements. Therefore, carrier frequencies L1 and L2 has short wavelengths. For both methods, the total error does not exceed 1–2 mm. This method requires the determination of the number of full wavelengths (n) traveled in the distance between the satellite and the receiver plus the fraction of a wavelength $\Delta\lambda$ corresponding to $\Delta\varphi$, where

$$\Delta\lambda = \lambda\Delta\varphi/2\pi .$$

Thus, $S = n\lambda + \lambda\Delta\varphi/2\pi$.

They are defined by the number of full wavelengths n that are traveled within the measured distance. In fact, only a fraction of the wavelength is determined for the measurement, as illustrated in Fig. 11.24. However, the problem of disambiguation, that is determination n remains. n must be defined. n-full wavelengths that pass within measured distance. Definition n or the disambiguation of the phase measurements is solved by observing the signals from the satellite over a long period, by using the two frequencies L1 and L2, and by using a combination of the phase and code methods. Considering a wavelength of 19 cm and the typical altitude of a satellite (20 180 km), a typical value of n is 100 000 000. However, there is an additional problem. When the signal is received, the clocks onboard the GNSS satellites and the clock of the receiver are not in sync. Therefore, a so-called *pseudodistance* is computed rather than the geometrically correct distance. GPS receives a pseudodistance. Eliminating the difference from the real distance requires the use of at least four satellites. Measurements of distances from satellites to the receiver are accompanied by errors determing the travel time of the signal from the satellite to the receiver. Since there are time measurement errors, the term pseudodistance (pseude-range) is used for such distances. Eliminating the difference from the real distance requires the use of at least four satellites. However, the PDOP (positional dilution of precision) is important too. The PDOP causes the accuracy of the position pinpointed by a GNSS to decrease to 12 m. The accuracy of measurements largely depends on the mutual location of satellites relative to the receiver antenna on the earth. When satellites in the scope are too close to each other, they talk about "poor" location geometry. PDOP is a numerical characteristic of location geometry. PDOP can reduce the accuracy of the GNSS position to 12 m.

11.5.2 Differential GNSS

A number of errors arise during measurements. Random errors are eliminated by averaging a sufficiently large set of measurements. There are also the following sources of systematic error:

- Errors in the basic data. These include inaccuracies in the ephemerides of the satellites that define the positions in space.
- Errors caused by external influences on the propagation of a range signal, such as ionospheric and tropospheric delays, reflections of radio waves from nearby objects (a multipath), obstacles, and relativistic effects.
- Instability in the operation of the signal generator, which leads to signal delays in the satellite and the receiver.
- Inaccurately defined position of the phase center of the receiver's antenna.
- Geometrical factors: the influence of the geometry of the relative positions of the observed satellites on the measurements.

There are also other error sources that are not related to the groups listed above. An example includes the error resulting from the transformation of coordinates from the geocentric coordinate reference system to a local system.

The relative path of a measurement is used to eliminate the errors caused by the influence of the atmosphere (Fig. 11.24).

In this case, measurements of two points A and B are obtained using the same satellites. The geometric difference between the coordinates of the points is given by the measurements. Both points are observed at the same time. For this reason, the influence of the atmosphere on the determination of the coordinates of A is identical to its influence on the determination of the coordinates of B, so the influence of the atmosphere on the coordinates of ΔX_{AB} and ΔY_{AB} can be excluded.

The differential method is applied to increase the accuracy of the measurements of a point on the Earth's surface or in near-Earth space. The method improves operational parameters such as accuracy, reliability, and availability, and the calculations involved utilize external data. The name of this method is often abbreviated to DGNSS (differential global navigation satellite system). A satellite system that uses differential corrections can provide coordinates within a territory up to 30 km in diameter to centimeter accuracy. Today, three satellite systems broadcast a differential signal free of charge. Such systems are collectively known as space-based augmentation systems (SBASs). The first of these systems is WAAS (Wide Area Augmentation System), which covers the territory of North America. The second system is EGNOS (European Geostationary Navigation Overlay Services), which covers the territory of Europe. Finally, MSAS covers the territory of Japan and some of its neighboring countries.

The most important relative measurement methods are as follows:

- In the *static mode*, simultaneous measurements are performed by motionless receivers at two or several points. One of the receivers is defined as basic (i.e., it sits at the base point and is called the rover). The position(s) of the other receiver(s) is/are defined relative to this point. Usually, measurements in static mode are carried out at long distances (i.e., longer than 15 km). The time taken for an observation depends on the distance between the points, the number of available satellites, the conditions in the ionosphere and the troposphere, and the accuracy demanded, but is usually less than 1 h.
- Using the *fast static mode* allows the duration of measurements to be reduced, because disambiguation algorithms are used. The durations of observations performed in this mode are 5–20 min.
  The relative accuracy achievable with static relative positioning is $\pm(3\text{–}5\,\mathrm{mm} + 0.5\,\mathrm{ppm})$ [2].
- The *kinematic mode* allows the determination of the coordinates of mobile stations while they are moving. In this mode, the receivers at the basic and mobile stations must stay in continuous contact with the satellites without interruption. Before the mobile stations begin to move, it is necessary to make phase measurements at a point with known coordinates.
  The relative accuracy achievable with kinematic relative positioning is $\pm(8\text{–}15\,\mathrm{mm} + 1\,\mathrm{ppm})$ [2].
- The *stop-and-go mode* is a specific kind of kinematic mode where the mobile station (the rover) stops moving before each point measurement and remains motionless for 5–30 s. Usually, measurements are repeated in several epochs to increase accuracy.
- In the *RTK (real-time kinematic)* mode, coordinates are obtained with centimeter accuracy by operating two stations: a first receiver—the base station—at a point with known coordinates, and a second receiver (the rover), which is used to perform measurements at new points. The offset between the known coordinates at the base station and its measured coordinates is computed. This offset is mostly caused by unaccounted-for atmospheric influences and geometric distortions of the local coordinate reference system. Assuming that those influences are identical at the base station and at the rover's position, the measured coordinates of the rover can be corrected by applying the offset vector with an opposite sign. This method works well if the distance between the receivers is no more than 10–30 km. The correction values can be transmitted to the rover online (immediate correction) or offline (postprocessing). Immediate correction requires a radio modem between both stations. The transfer speed may be 2400 bps or more, which causes a transfer delay of no more than 0.5–2 s. However, a speed of only 200 bps, leading to a delay of 10 s, is more typical, and is still sufficient for practical applications. The main advantage of RTK is the possibility of exact processing in real time. Usually, amendments are transmitted in RTCM SC-104 format. RTCM version 10 403.3, available since October 2016, works with GPS, GLONASS, Galileo, QZSS (Quasi-Zenith Satellite System, Japan), and BeiDou.

As shown, there are several methods of applying navigation amendments: postprocessing, DGNSS, and RTK. They differ in the accuracy of the final result and in the time spent applying the corrections. The postprocessing mode achieves the highest accuracy but demands considerable time for collection and data processing. In the DGNSS mode, the time spent is significantly less. Indeed, the amendments can be performed in real time. However, the accuracy of the amendments made in the DGNSS mode is limited to about a meter. The RTK mode also allows amendment in real time, and yields an accuracy of about 1 cm in x and y and 2 cm in height. Therefore, the RTK method allows surveys to be carried out using only satellite receivers. Because of their similar levels of accuracy, RTK can be used in combination with a total station.

11.6 More Surveying Methods

11.6.1 Traversing

A traverse is a sequence of straight lines whose ends are marked and whose lengths, as well as the angles between them, are subject to observations. A traverse used to be the most widespread method of determining the relative locations of points.

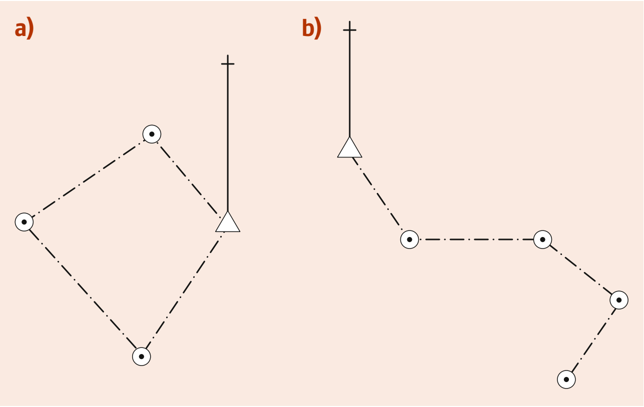


Fig. 11.25 **a** Closed traverse (polygon). **b** Open traverse

Traverses may be closed or open. There are two kinds of closed traverses: polygon and link. Figure 11.25a depicts a polygon. An open traverse is shown in Fig. 11.25b.

A closed traverse allows the quality of measurements to be assessed, because the actual coordinates of the endpoint can be compared with its nominal coordinates. In the case of a polygon, the nominal coordinates are the coordinates of the first point. In the case of a link, the nominal coordinates are the coordinates of the endpoint. In a similar way, the actual value and the nominal value of the reference direction can be compared.

If the comparison shows a gross error, the measurements shall be monitored and eventually repeated. If the remaining errors are within the accuracy of the instruments, then the resulting inconsistencies in coordinates and angles are distributed according to the least squares method (strong method) or proportional to line length (approximate method).

The hanging travers rests on only one point with known coordinates. There is not the second point with known coordinates.

11.6.2 Point Measurement

Often, only the coordinates of a single point are needed. GNSS measurements aside, there are three main methods of achieving this: intersection of lines, 3-point resection, and 2-point resection.

The intersection of lines method, also called the direction–direction problem, requires one distance and two angle measurements or, if the positions of the stations are known, just the two angle measurements. The result is the position of the (distant) point of interest (Fig. 11.26a).

3-Point resection requires three distant reference points with known positions and a theoretical minimum of two angle measurements between those points. However, a third (redundant) angle is always measured as well in order to check the measurement. The result is the position of the station where the angle measurements were taken (Fig. 11.26b).

2-Point resection requires two distant reference points with known positions, two distance measurements from the point of interest to the two given points, and a measurement of the angle between the two directions. The result is the position of the station where the angle measurements were taken (Fig. 11.26c).

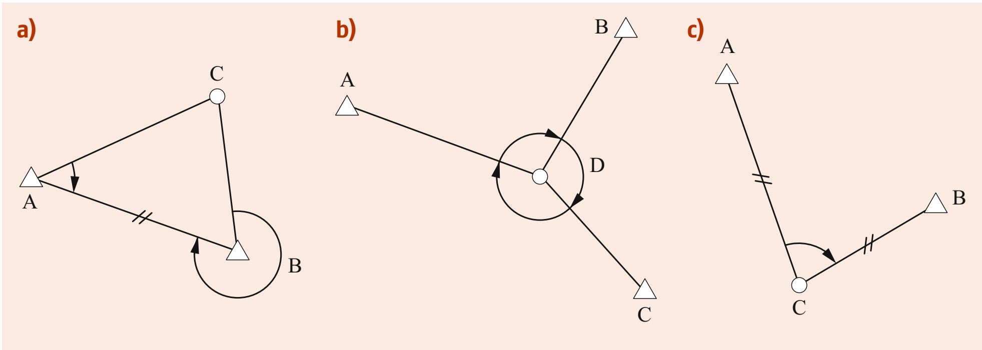
11.6.3 Network Measurement

Triangulation

A triangulation survey consists of a network of triangles in which one side length and all the angles are measured. The lengths of all the other sides are computed without further measurement. The single measured line is the base line of the network; this defines the scale of the network and must be measured with very high accuracy. The measured and computed lengths of a check base line at the end of the network can be compared to check the work [4]. Triangulation was practically the only method available for surveying large areas before the invention of electronic distance measurement in the middle of the twentieth century (see Fig. 11.27).

The base line length is usually much less than the typical side length of a triangle in the network, and the base is extended by triangulation until one of the main sides of the triangle can be computed. In the example shown in Fig. 11.28, the measured base is ab. This is extended to c

Fig. 11.26 **a** Intersection of lines method (direction–direction problem), **b** 3-point resection, **c** 2-point resection



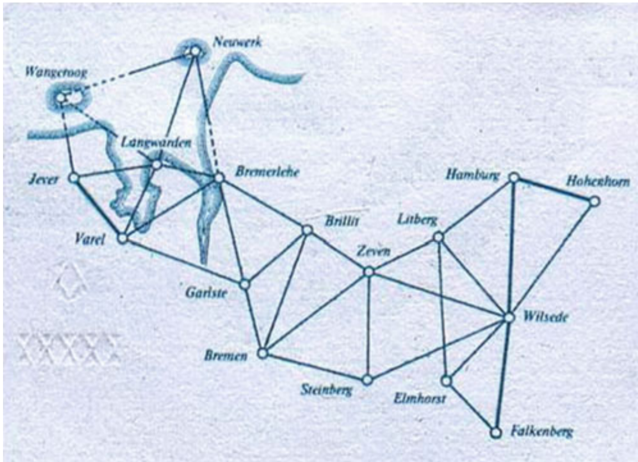


Fig. 11.27 Triangulation of the northern part of the Kingdom of Hanover (now part of Germany) by Carl Friedrich Gauss, 1823–1825, as shown on the 10-Deutsche Mark bill formerly used as legal tender in Germany

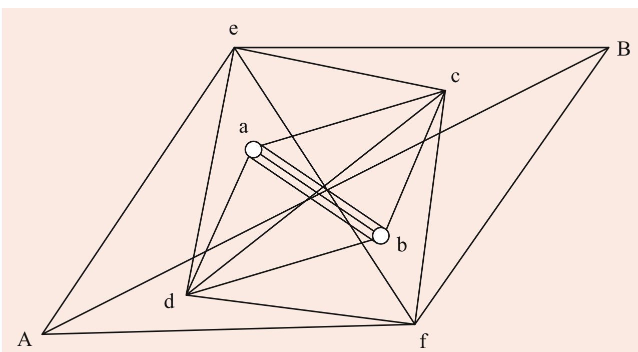


Fig. 11.28 Network used for base line measurement

and d, and then to e and f, and finally to the triangulation stations A and B [4].

Trilateration

This method involves setting up networks of triangles similar to those shown above for triangulation, but measuring all the side lengths rather than the angles. This technique has been made possible by the development of electronic distance measurement (EDM).

References

1. . https://www.fig.net/about/general/definition/definition.pdf, Accessed 8 Feb 2019
2. Ghilani, C.D.: Elementary Surveying – An Introduction to Geomatics, 15th edn. Pearson, New York (2018)
3. Jäger, R.: GOCA – GNSS/LPS/LS-based online control and alarm system (2019). http://goca.info/, Accessed 2 Jan 2019
4. Whyte, W.S., Paul, R.E.: Basic Surveying, 4th edn. Taylor & Francis, London (1997)
5. Nadolinets, L., Levin, E., Akhmedov, D.: Surveying Instruments and Technology. Taylor & Francis, CRC Press, Boca Raton (2017)
6. Golubev, V.: Theory of Mathematical Processing of Geodetic Measurements. MIIGAiK, Moscow (2016)

Further Reading

7. Aggarwal, Priyanka, Syed, Zainab, Noureldin, Aboelmagd, El-Sheimy, Naser: MEMS Based Integrated Navigation. Artech House, Norwood (2010)
8. Agor, R.: A Textbook of Surveying and Levelling, 12th edn. Khanna, Delhi (1980)
9. Deumlich, F.: Surveying Instruments. De Gruyter, Berlin, New York (1982)
10. Doucet, A., de Freitas, N., Gordon, N.: Sequential Monte Carlo Methods in Practice. Springer, Heidelberg (2010)
11. Grewal, M.S., Andrews, A.P., Bartone, C.G.: Global Navigation Satellite Systems, Inertial Navigation and Integration, 3rd edn. Wiley, Hoboken (2020)
12. Hofmann-Wellenhof, B., Lichtenegger, H., Wasle, E.: GNSS – Global Navigation Satellite Systems. Springer, Heidelberg (2008)
13. Hooijberg, M.: Practical Geodesy. Springer, Heidelberg (1997)
14. Jekeli, C.: Inertial Navigation Systems with Geodetic Applications. De Gruyter, Berlin, New York (2001)
15. Johnson, A.: Plane and Geodetic Surveying – The Management of Control Networks. Spon, Taylor and Francis, London, New York (2004)
16. Kavanagh, B.F.: Surveying with Construction Applications, 7th edn. Prentice Hall, NJ, USA (2010)
17. Leick, A., Rapoport, L., Tatarnikov, D.: GPS Satellite Surveying, 4th edn. Wiley, Hoboken (2015). Kindle Edition
18. Roy, S.K.: Fundamentals of Surveying. Asoke K. Ghosh, Prentice-hall of India, New-Delhi (1997)
19. Rüeger, J.M.: Electronic Distance Measurement. Springer, Heidelberg (1996)
20. Seeber, G.: Satellite Geodesy. De Gruyter, Berlin, New York (2003)
21. Torge, W., Müller, J.: Geodesy, 4th edn. De Gruyter, Berlin, New York (2012)

Vladimir Golubev Vladimir Golubev is Professor at the Moscow University of Geodesy and Cartography and is heavily involved in the development, advancement, and performance of surveying and geodesy training. For his work of nearly 50 years in the field, more than 20 of which as a Head of the Faculty of Territory Development, he was awarded with the title "Honorary Surveyor of Russia".

12 Geometry and Topology

Gerhard Gröger and Betsy George

Contents

Abstract

The representation of the geometrical properties of spatial objects and their structural aspects (topology) is crucial for GIS operations, analyses, and visualizations. This chapter introduces the most important geometrical and topological concepts, considering the two-dimensional as well as the three-dimensional case. Particularly, the concepts of the standard ISO 19107 Spatial schema are introduced.

Keywords

Geography Markup Language · Geometry · 3-D Geometry · Topology · Boundary Representation · Topological Relations · Topological Data Models

G. Gröger (✉)
CPA ReDev GmbH
Siegburg, Germany
e-mail: groeger@uni-bonn.de

B. George
Spatial and Location Technologies, Oracle America, Inc.
Nashua, NH, USA
e-mail: betsy.george@oracle.com

12.1 Geometry

The main purpose of geometrical characterizations in GIS is to represent the shape and metric properties of spatial objects (features) in order to compute distances between objects, to derive areas of surface objects or volumes of solid objects, to perform spatial analyses (Sect. 6.2.8) like viewshed calculation, spatial planning and simulation, e.g., for noise emission, or to serve as base for geovisualization (Chaps. 11 and 14).

Regarding the dimension of geometry, we have to distinguish between the *dimension of the geometry object* and the *dimension of the embedding space*. In this section, the embedding space is, in general, 3-D; embedding in 2-D planes are considered only in Sect. 12.1.2. This section is structured according to the dimension of the geometry objects: we start with 0-D to 2-D, consider 2.5-D as special case, and finally discuss solid 3-D objects.

An overview of geometrical 3-D models can be found in [1–3]. This section presents models that are relevant for 3-D-GIS, mainly boundary representations (Sect. 12.1.3), and gives a rough survey of other concepts. The focus is on the geometry model provided by the standard ISO 19107 *Spatial schema* ([4] or Sect. 13.5), which is implemented particularly in the representation and exchange language *Geography Markup Language (GML 3)* ([5–7] or Sect. 15.5).

The 2-D coordinates $(x, y) \in \mathbb{R}^2$ or 3-D coordinates $(x, y, z) \in \mathbb{R}^3$ of the geometry objects are represented according to any of the coordinate reference systems introduced in Sect. 12.2.2.

The standard ISO 19107 distinguishes between *geometric primitives* and aggregated geometric objects (collections), which are composed of geometric primitives. Primitives are presented first.

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_12

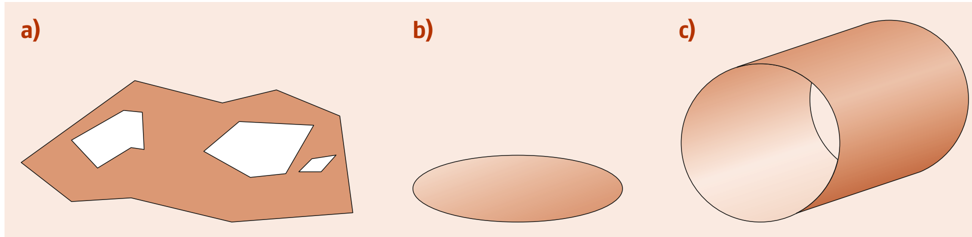


Fig. 12.1 Three surfaces. **(a)** Planar polygon delimited by four rings. **(b)** 2-D disk delimited by one ring. **(c)** Cylinder surface delimited by two rings

12.1.1 0-D, 1-D, and 2-D Geometries

A *point* as a 0-D geometry is simply represented by a 3-D coordinate (x, y, z). One-dimensional geometries are *curves* or *line segments*, which have start coordinates and end coordinates. The shape of a curve between the start point and the end point is specified by an *interpolation method*. The list of interpolation methods provided by the ISO 19136-x GML, for example, is *linear*, *geodesic*, *circular*, *elliptical*, *clothoid*, *conic*, *polynomialSpline*, *cubicSpline*, or *rationalSpline*. If the interpolation is linear, start and end points are connected by a straight line. The other interpolation methods require some more parameter values. A circular line segment is represented by three control points, and an elliptical one by four control points, for example. More details on interpolation methods are provided in [3, 4].

In general, curve primitives (*class Curve*) are connected, nonbranching (i.e., have at most two start/end points), and are non self-intersecting. If the positions of the start and end points are identical, the curve is *closed*.

2-D geometric primitives embedded in 3-D space, which are typically called *regions*, *polygons*, or *surfaces* (class *GM_Surface*), are continuous, connected 2-D point sets that are delimited by curves. These curves have to be closed and form so-called *rings* [4]. A *ring* is a closed sequence of curves, where a curve starts where the predecessor in the curve ends. The curves in a ring are nonintersecting. A region is bounded by one exterior ring and by optional interior rings, which define enclaves or holes in the region. Figure 12.1a depicts as an example a region with four rings, one exterior, and three interior. Rings may be composed of curves of any interpolation method mentioned in this section.

The shape of the surface, i.e., of the 2-D point set delimited by its rings, is defined by interpolation methods, similar to the case of line segments. ISO 19107 *Spatial schema* and ISO 19136-x GML, for example, provides the following interpolation methods: *planar*, *spherical*, *elliptical*, *conic*, *tin*, *parametricCurve*, *polynomialSpline*, *rationalSpline*, *triangulatedSpline*. In planar surfaces, all points of the surface are in the same plane. This interpolation method is common in GIS and 3-D city models; surfaces provided by commercial tools like ArcGIS or Oracle Spatial (Sect. 3.11) are planar. For details of the representations of other interpolation methods, the reader is referred to [3, 4].

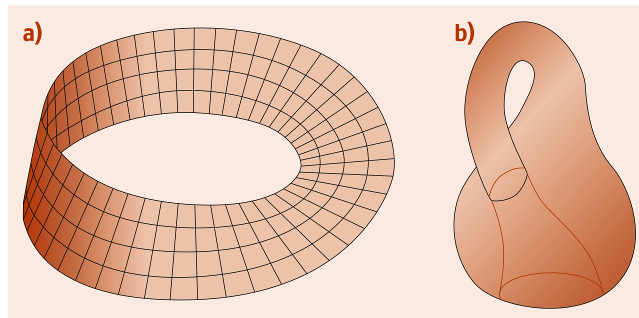


Fig. 12.2 Nonorientable surfaces. **(a)** Möbius strip, **(b)** Klein bottle

Surfaces are purely areal two-dimensional point sets, without penetrations, T-shaped, or X-shaped touches. Mathematically, this property is captured by the notion of a *2-manifold*. A *2-manifold* is a 2-D point set where each point has a neighborhood in the set that is topologically equivalent to an open two-dimensional disk. Intuitively, for each point on a 2-manifold, a small circular neighborhood centered at that point can be deformed to a disk.

Another important property of surfaces is the number of boundaries. A disk (Fig. 12.1b) has one boundary, whereas the number of boundaries of a cylinder surface (Fig. 12.1c) is two. A sphere has no boundaries at all. Such surfaces are called *closed*; they enclose a volume completely and, hence, are used to define solids (Sect. 12.1.3).

A further relevant characteristic of surfaces embedded in 3-D space and an essential precondition for defining solid objects is *orientability* [2]. A surface is *orientable*, if two opposite sides of the surface can be distinguished. For the general case, a more formal definition of orientability is given in [8]. Well-known examples for nonorientable surfaces are the Möbius strip and the Klein bottle; both are depicted in Fig. 12.2.

An orientable surface can be given an orientation, by labeling exactly one of the two sides as *top* or $+$. When surfaces are used to define solids (Sect. 2.1.3), the surface orientation is chosen such that the top side points outward relative to the solid's interior. If one of the rings delimiting a surface can be distinguished as an outer ring, as is the case for planar polygons, the *right-hand rule* can be applied to define an orientation. If the curve segments in the exterior ring are oriented consistently, from start point to end

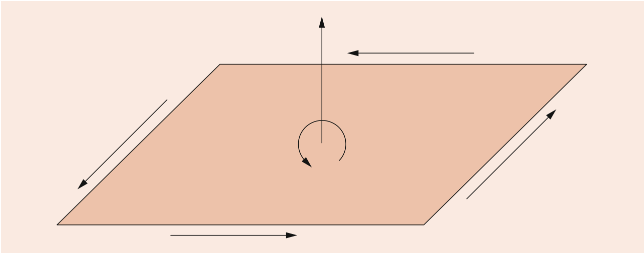


Fig. 12.3 Orientation of a planar surface by applying the right-hand rule

point, then the side of the surface where the direction of the ring appears counterclockwise is the top side of the surface (Fig. 12.3).

The surface and curve primitives can be aggregated to larger units, which again have the same properties as the surface or curve. Such objects are called *composites*. Composite surfaces (curves) can recursively again be part of a larger composite surface (curve). Rules for building composite surfaces or curves from parts are discussed in Sect. 12.2.2, where topological data models are reviewed. In fact, a composite surface (curve) is both a topological cell complex as well as a surface (curve).

There is also an aggregation concept below the level of primitives. *Line segments* are aggregated to primitive curves, and *surface patches* are aggregated to primitive surfaces.

In ISO 19107 and in ISO 19136-x GML, there is another type of aggregation of primitives, which is defined less strictly and does not provide the properties that composites have. This type is called *collection* in ISO 19107 or *aggregate* in GML. In GML, its subtype for 2-D objects is called a *multisurface*. The surfaces that are part of a multisurface may overlap or penetrate, and the aggregates may be unconnected. A similar concept is defined for curves. The corresponding subtype of aggregates is called a *multicurve* in GML, which may be unconnected and branching, and the curves in a multicurve may intersect.

12.1.2 Special Cases: 2-D as Embedding Space and 2.5-D

In most commercial 2-D GIS, geometries are embedded in 2-D space, i.e., the third coordinate is omitted or set to zero. Obviously, the interpolation of regions has to be planar. An example for a 2-D geometry model is *GML2* [5].

In the so-called *2.5-D geometry model*, the embedding space is 3-D, and the geometries are 0-D to 2-D. However, but there is an important restriction: for each geometry, the height value z is a function of each x/y-point, i.e., at each x/y-point, there is at most one height value. Typically, a 2.5-D model is used to represent the terrain surface; in that case it is called a *digital terrain model* (Sect. 9.5.1). Due to the functional dependency between the planar and the height values, vertical walls and overhangs, e.g., the wall of a building or a balcony, are outside the scope of a 2.5-D model.

12.1.3 3-D Geometries

Spatial objects like buildings, rooms, or other volume objects are represented by *solids*. In geometrical modeling, solids are described mathematically by *rigid bodies* [9]. A rigid body is a bounded, regular, and semianalytical subset of $\mathbb{R}^3$. Regularity excludes nonvolume elements like point or line enclaves, while semianalytical sets are constructed by combining analytical sets – which are the range of analytical functions, particularly polynomials – by the set operations difference, intersection, and union. In boundary representation schemes, which are widely used in geometrical modeling, CAD, and GIS, solids are represented by their bounding surfaces. Rigid bodies are exactly those bodies that are bounded by a single, closed 2-manifold [9].

In geometrical modeling, there are different schemas to represent solids. The most important are reviewed in the following sections: the boundary representation, constructive solid geometry, raster-based or enumeration methods, sweep representations, and primitive instancing. For a comparison of these representations on the basis of several criteria (accuracy, domain, uniqueness, validity, closure, compactness, and efficiency) the reader is referred to *Foley* et al. [1].

Boundary Representations

The most common representation schema for solids in GIS is the *boundary representation* [1–4], which defines solids by its bounding surfaces. The saddle-roof building in Fig. 12.4, for example, is modeled by a solid, which is bounded by seven planar surfaces: a ground surface, four wall surfaces, and two roof surfaces.

The composite surface or the set of surfaces bounding a solid must obey the following three conditions.

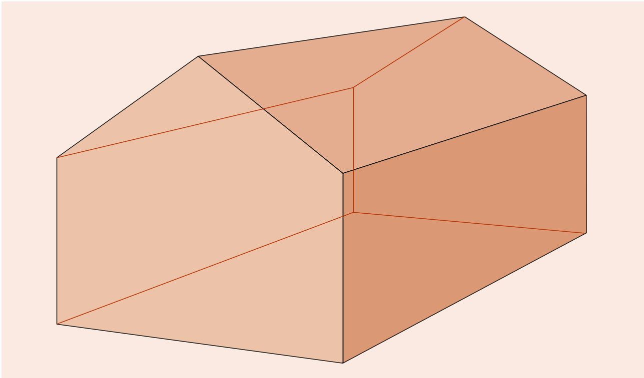


Fig. 12.4 Boundary representation of a saddle-roof building

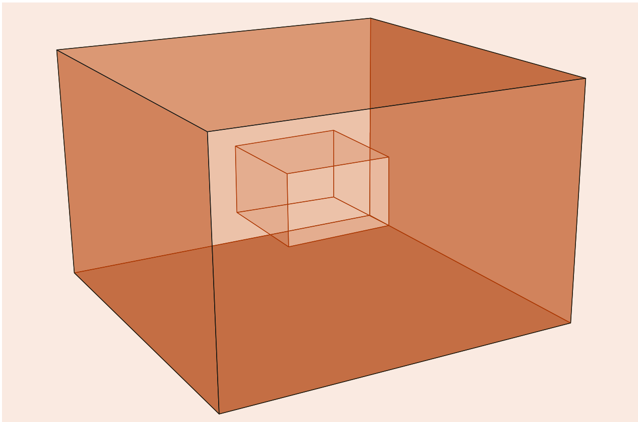


Fig. 12.5 Solid bounded by one exterior shell and one interior shell forming an enclave

1. The surfaces have to enclose the solid completely, without gaps. This requirement is particularly met by closed (composite) surfaces.
2. The surfaces have to be purely areal and nonoverlapping, i.e., must be a 2-manifold.
3. The surfaces have to be orientable and must be oriented in such a way that the top side of all surfaces points outward from the solid's interior (Sect. 12.1.1).

However, each closed surface embedded in 3-D space without penetrations is orientable. This important theorem is implied by a well-known proposition for closed surfaces, which states that each closed surface embedded in 3-D space without penetrations is homeomorphic to a sphere with $n \geq 0$ handles, which is orientable [3, 9]. Hence, orientability does not need to be checked.

The solids provided by the standard ISO 19107 may have enclaves, which define volume voids inside the solid. To represent enclaves, the surfaces bounding a solid are grouped in so-called *shells*. Each solid is delimited by exactly one exterior shell (which has already been defined by the three conditions above) and optional interior shells, each bounding a void (Fig. 12.5). The interior shells have to fulfill the three conditions given above; particularly, the top sides of all surfaces defining the shell have to point towards the enclave.

The aggregation concepts that were already introduced for curves and surfaces are provided for solids as well. A *composite solid* is a topological cell complex consisting of solids, which is again a solid, i.e., the exterior and all interior shells of which each fulfills the three conditions given above. A *GM_MultiSolid* is an aggregation of solids, which does not obey any restrictions, i.e., which may penetrate each other or which may be unconnected.

Some models define special solids that are not bounded by a shell completely; such solids are used to define 3-D tessellations, i.e., the complete coverage of 3-D space by solids.

Constructive Solid Geometry (CSG)

Constructive solid geometry (CSG) [2, 3] is a procedural modeling technique that allows us to create solid objects by using *Boolean operators* to combine primitive objects. As primitives, boxes or cylinders are typically used. Operations are transformations (rotation, scaling, and translation) and set operations (union, intersection, and difference). Figure 12.6 depicts, as an example, the representation of a solid saddle-roof building as a CSG tree. Figure 12.6 gives an example of a representation as a solid saddle-roof building as a CSG tree. From one box, another box is subtracted (operator /) after transformation (translation and rotation), and from the resulting box, yet another box is subtracted, also after transformation (translation and rotation). To avoid generating nonsolid parts, *regularized* Boolean operators are employed, which remove all purely areal, linear, or point objects.

CSG is an implicit representation; not the result of the derivation is represented but the CSG tree containing the sequence of operations and the corresponding parameters and primitives.

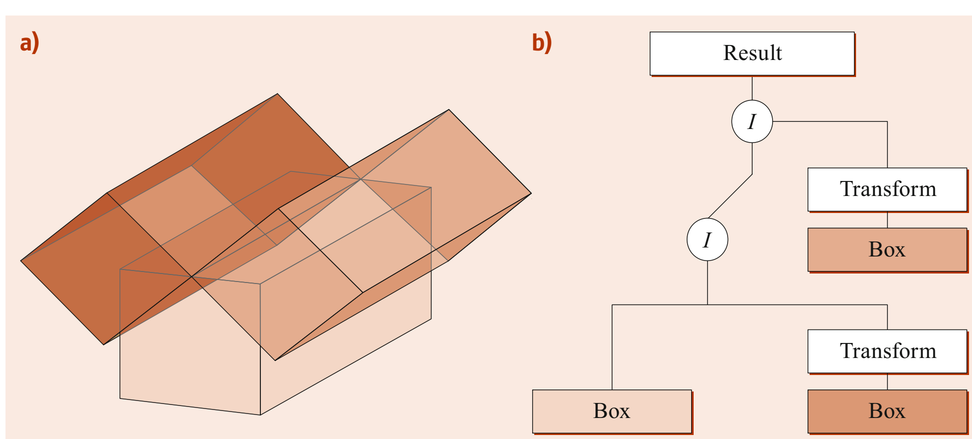


Fig. 12.6 CSG model of a simple saddle-roof building. **(a)** Transformed primitives, **(b)** CSG tree

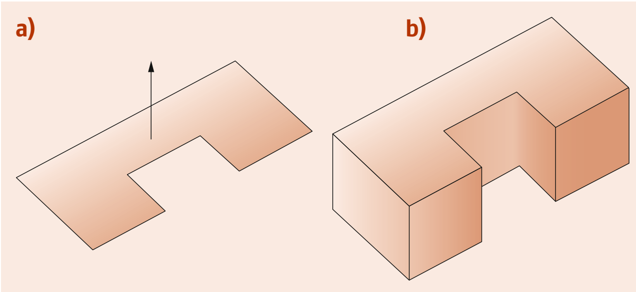


Fig. 12.7 Sweeping a polygon yielding a solid. (Adapted from [2])

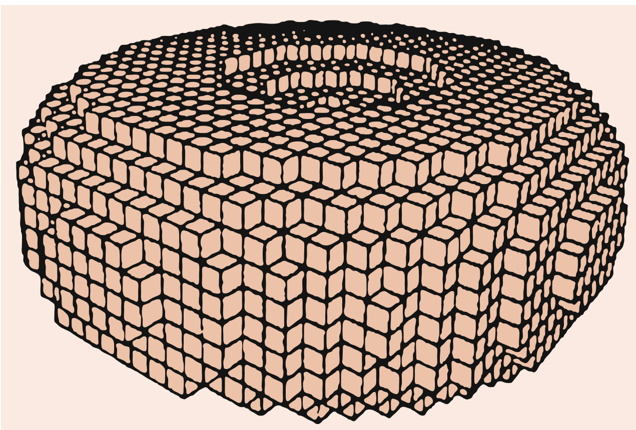


Fig. 12.8 Representation of a solid by voxels. (Adapted from [2])

Sweep Representations

In *sweep representations*, a solid is generated by sweeping a surface along a curve. All points in space that are touched by sweeping the surface constitute the solid. Three kinds of curves are typically used: lines (translation sweep), circles (rotation sweep), and combinations of both.

A special case of sweeping is *extrusion*, where a planar, horizontal polygon is swept along a line that is perpendicular to that surface (Fig. 12.7). This method is often used to construct buildings in a 3-D city model in the less detailed level of detail 1 (blocks models, [10]) from cadastral footprints. However, extrusion in that case is used as the method for constructing the model; for storing the result of the extrusion, mostly a boundary representation is used.

Raster-Based/Enumeration Methods

In *enumeration methods* [2], space is partitioned into regular cells (cuboids), which are called *voxels*. A solid, for example, can be represented by listing all voxels contained (completely or partially) in the solid (Fig. 12.8). This method generalizes the partitioning of 2-D planes in rectangular raster cells.

Primitive Instancing

The modeling schema of *primitive instancing* [1, 2] enables the representation of predefined, parameterized geometrical primitive object types. To yield a geometrical instance of that type, the variable parameters have to be instantiated. An example is depicted in Fig. 12.9, where a t-brick primitive with five parameters w_1, w_2, h_1, h_2, and l is constructed by instantiating the corresponding parameter values. An important application of primitive instancing in GIS is model-based building classification and 3-D reconstruction from aerial laser scanning or imagery. The predefined primitive object types are roof types (flat, monopitch, saddle, hipped). A saddle-roof type, for example, has four parameters: ridge height, eaves height, length, and width, under the assumption that the building is symmetrical.

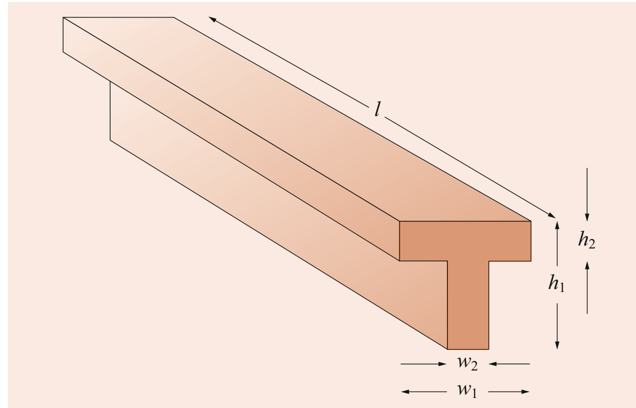


Fig. 12.9 t-brick constructed by primitive instancing. (Adapted from [2])

12.2 Topology

Whereas geometry deals with the shape and metric properties of spatial objects, *topology* focuses on the structure of and qualitative relations between spatial objects, i.e., whether two objects overlap or whether one object is contained in another. In mathematical topology, two branches relevant for 3-D-GIS are differentiated between: *point-set topology* and *algebraic topology*. Point-set topology aims at defining and classifying qualitative relations between spatial objects. These relations, also called *topological predicates*, provide essential elements of *spatial query languages*, e.g., to formalize a query retrieving all municipalities in North Rhine-Westphalia or to obtain all highways passing through California. Topological predicates are used in *OGC Filter encoding* (ISO 19143, [11] or Sect. 15.5.4), employed in the context of spatial data infrastructures, in the query language of the commercial database *Oracle Spatial* ([12] or Sect. 3.11), in query languages provided by the commercial GIS ArcGIS, and in standards like ISO 19107 [4].

Algebraic topology [13, 14] is the basis of many data models in GIS, CAD, and computer graphics, which are called *topological data models*. This branch of mathematics provides formal rules to construct complex objects from primitive ones, avoiding penetrations and modeling touches of spatial objects explicitly. Topological data models aim at providing efficient navigational access without considering geometry and serve as a basis for the definition of consistent models. In this section, we first discuss topological relations in different dimensions and then address 3-D topological data models.

12.2.1 Topological Relations

In general, the specification of topological relations is defined for arbitrary *topological spaces*. A *topological space* [8], which is a fundamental notion in topology, is a set M together with a set N of subsets of M, called *neighborhoods*, where the following conditions hold.

1. Each element $m \in M$ is in a neighborhood $n \in N$.
2. The intersection of two neighborhoods of $m \in M$ is or contains a neighborhood of m.

Let M be a topological space and X a subset of M. An element $p \in M$ is *near* X, if each neighborhood of p contains an element in X. The *interior* of X, denoted X°, is the set of all elements in X, which are not elements near the complement of X. The *boundary* of X, denoted ∂X, is the set of all elements that are both near X and the complement of X. The *exterior* X^- of X is the complement of the union of the boundary and the interior.

To illustrate these concepts, let M be the set $\mathbb{R}^3$ and the neighborhoods $n \in N$ be defined by open balls. Then the interior and the boundary of spatial objects (point sets) has an intuitive meaning; X may, for example, be a box. The interior of the box (X°) is the point set bounded by the six rectangles defining the box, and the boundary (∂X) is defined by the six rectangles. The exterior of the box (X^-) is the space outside the box.

Topological relations can be defined based purely on the notions of interior, boundary, and exterior. We now focus on the 4-intersection and the 9-intersection model and its extensions in 2-D and 3-D. Another similar approach for defining topological relations is region connection calculus [15].

2-D

The well-known *4-intersection model* introduced by *Egenhofer* and *Franzosa* [16] defines binary topological relations in 2-D, i.e., relations between two objects. For two regions A and B, the intersections ($\cap$) of the interiors (A°, B°) and the boundaries (∂A, ∂B) are considered systematically, in which it is only relevant whether or not the intersection is empty.

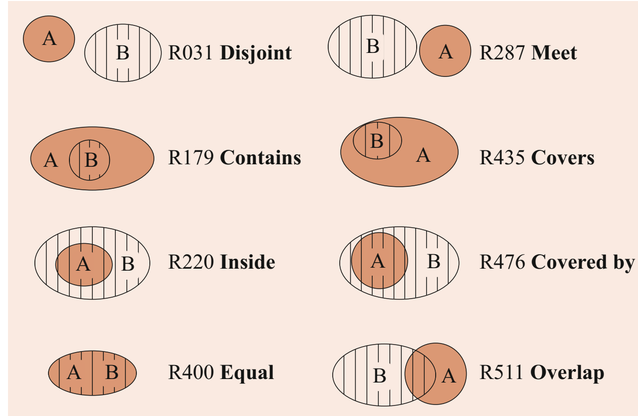


Fig. 12.10 Eight relations for simple regions distinguishable by Egenhofer's 4-intersection model. (Adapted from [16])

The result is represented by a Boolean 2×2 matrix (an empty intersection $\varnothing$ is denoted by *false*, a nonempty $\varnothing$ by *true*)

$$\begin{vmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B \\ \partial A \cap B^\circ & \partial A \cap \partial B \end{vmatrix}.$$

The regions considered in that model are restricted topologically; a region must be bounded by a single connected, closed curve, which is non self-intersecting (i.e., a *ring* as defined in Sect. 12.1.3) and hence, does not have holes. Not all of the $2^4 = 32$ relations have a spatial realization in that model; only eight relations are possible (Fig. 12.10). The other 24 relations cannot occur due to dependencies between the values of matrix elements. For example, if a boundary–interior or interior–boundary intersection is non-empty, then the interior–interior intersection is non-empty as well.

This model can also be applied to points and curves, where the boundary of a curve is defined as the union of both end points, the interior is the curve without the end points. The boundary of a point is empty, and the interior is the point itself. A line object must be connected, non-branching, and non self-intersecting.

Two extensions of the 4-intersection model have been developed, which both provide a more fine-grained differentiation of spatial arrangements. The first extension is to consider the *exterior* A^- of a point set A as well. In this *9-intersection model* [17], a 3×3 matrix denotes the Boolean intersection values. The relations of the 9-intersection model are also called *Egenhofer operators* [4]

$$\begin{vmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap B^- \\ \partial A \cap B^\circ & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^\circ & A^- \cap \partial B & A^- \cap B^- \end{vmatrix}.$$

Figure 12.11 depicts an example demonstrating that the 9-intersection model is more powerful than the 4-intersection

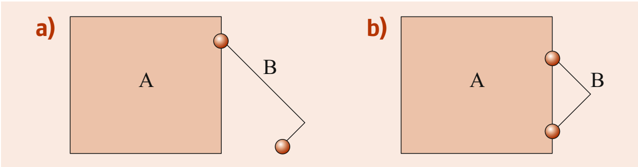


Fig. 12.11 Two topological different situations that can be distinguished by the 9-intersection model but not by the 4-intersection model

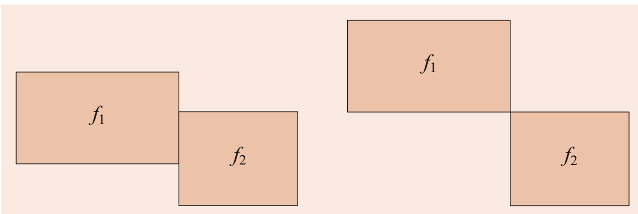


Fig. 12.12 Two situations that cannot be differentiated by the 4 or 9-intersection models but by considering the dimension of the intersection model. In Fig. 12.11a,b, the region A and a line segment B share boundaries, but in Fig. 12.11b, both enclose a region completely. Hence, both situations are different topologically. In the 4-intersection model, both situations are represented by the same relation: the intersection of the interiors and of the boundaries/interiors is empty, whereas the intersecting of the boundaries is not ($A^\circ \cap B^\circ = A^\circ \cap \partial B = \partial A \cap B^\circ = \varnothing$; $\partial A \cap \partial B = \neg\varnothing$). In the 9-intersection model, the intersection between the exterior of A and the boundary of B is nonempty in Fig. 12.11a ($A^- \cap \partial B = \neg\varnothing$), but empty in Fig. 12.11b ($A^- \cap \partial B = \varnothing$). Hence, both situations can be distinguished.

However, if 2-D objects embedded in 2-D space are considered (Fig. 12.10), the 4-intersection and the 9-intersection models yield identical results. In general, if the *codimension* – the difference between the dimension of the embedding space and the dimension of the objects – is zero, both models are identical.

A second extension of the 4-intersection model, which also can be applied to the 9-intersection model, is to consider the *dimension* of the intersection. The situations in Fig. 12.12 cannot be differentiated by the 4-intersection model or the 9-intersection model, since the relations *meet* in both cases, but both differ in the dimension of the intersection, which is 0-D or 1-D. The extension of the 4-intersection model to the 9-intersection model and the inclusion of the dimension of the intersection is called the *dimensionally extended 9-intersection model* or *DE-9IM* [18].

3-D Relations

Zlatanova [19] extended the 4-intersection model to 3-D by using $\mathbb{R}^3$ as the embedding space for points, lines, and surfaces, and by considering solids. Solids must have a single, connected, 2-manifold boundary. Figure 12.13 depicts all possible relations between two solids, which are identical to the relations between two surfaces in 2-D (Fig. 12.10).

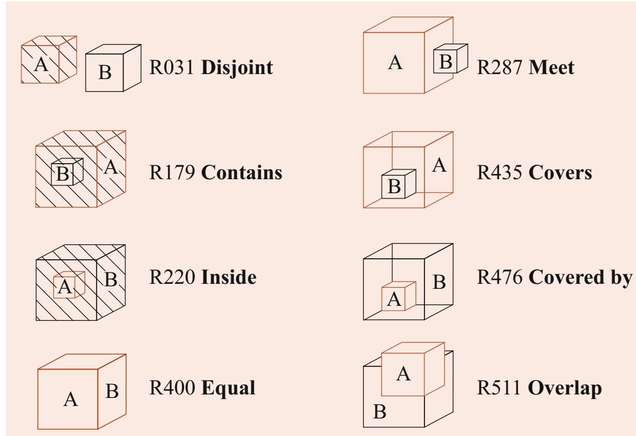


Fig. 12.13 Topological relations between two solids in $\mathbb{R}^3$. (Adapted from [19])

If the codimension is strictly greater than zero, a variety of topological relations is observed between two objects embedded in 3-D. As an example, Fig. 12.14 enumerates all 38 relations between two surfaces. These relations cannot be named meaningfully; they are denoted by a decimal code preceded by the character R, where the code is equivalent to the binary representation of the 9-intersection matrix. The order of the matrix elements in the binary code is as follows (the second row denotes the decimal and the third the binary representation of the summand corresponding to the relation in the first row)

$\partial A \cap \partial B$	$A^\circ \cap B^\circ$	$\partial A \cap B^\circ$
$2^8 = 256$	$2^7 = 128$	$2^6 = 64$
100000000	10000000	1000000
$A^\circ \cap \partial B$	$A^- \cap B^-$	$A^- \cap \partial B$
$2^5 = 32$	$2^4 = 16$	$2^3 = 8$
100000	10000	1000
$A^- \cap B^\circ$	$\partial A \cap B^-$	$A^\circ \cap B^-$
$2^2 = 4$	$2^1 = 2$	$2^0 = 1$
100	10	1

For example, the code R287 ($= 256 + 16 + 8 + 4 + 2 + 1$) is equivalent to the binary code 100011111 ($= 100000000 + 10000 + 1000 + 100 + 10 + 1$); in that relation, both boundaries intersect, and all other intersections not involving an exterior are empty ($\partial A \cap \partial B = A^- \cap B^- = A^- \cap \partial B = A^- \cap B^\circ = \partial A \cap B^- = A^\circ \cap B^- = \neg\varnothing$/true; $A^\circ \cap B^\circ = \partial A \cap B^\circ = A^\circ \cap \partial B = \varnothing$/false). The usage of the topological relations in query languages is by referencing the name of the relation. In *Oracle Spatial 11g*, a combination of relation names, connected by a logical *OR*, may be used. In the *simple features model* [20], ISO 19107 *Spatial schema* [4] and Or-

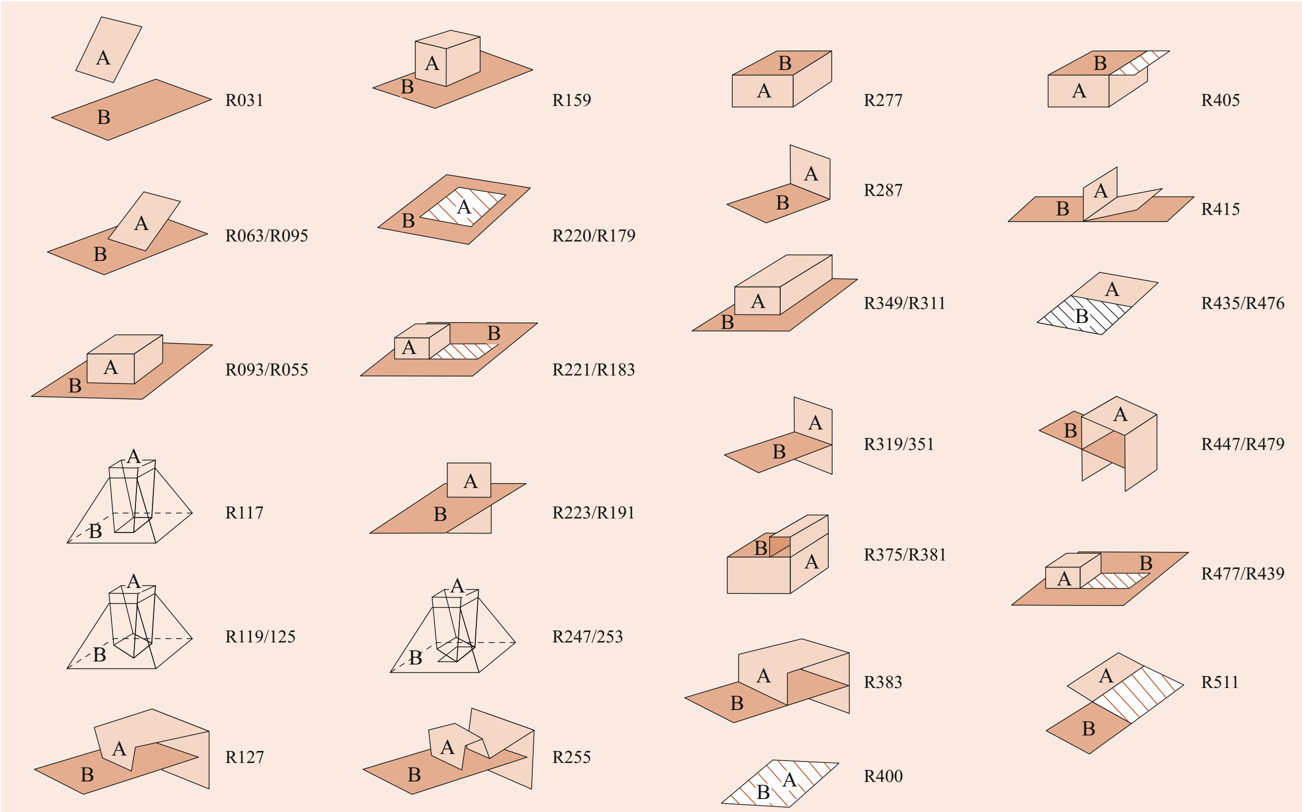


Fig. 12.14 Topological relations between two surfaces in $\mathbb{R}^3$. (Adapted from [19])

acle, a more flexible mechanism is also employed: a method *relate* receives the complete pattern of the 9-intersection matrix in row major form as input, containing the values F (empty intersection), the dimension 0, 1, 2, 3 of the intersection, or the wildcard symbol N (the value does not matter).

12.2.2 Topological Data Models

The theoretical foundation of all topological data models in GIS or CAD is the mathematical concept of *simplicial complexes* and its generalization, the concept of *cell complexes* [13, 14]. A cell complex consists of four types of primitives: 0-cells, also called nodes, *1-cell*, also called edges, *2-cells* (faces), and *3-cells* (topological solids). Each n-cell is topologically equivalent to a manifold of the corresponding dimension. For example, a 3-cell is topologically equivalent to a sphere, and a 2-cell to a 2-D-disk. Each n-cell c is bounded by $(n-1)$-cells $c_1, c_2, \ldots, c_k$, which are the *boundary* of the cell. Vice versa, c is in the *coboundary* of $c_1, \ldots, c_k$. A *cell complex* is an aggregation of n-cells, where the following condition holds.

- The intersection of two cells in the cell complex is either empty or a cell that is part of the boundaries of both cells.

Figure 12.15 gives two examples of cell complexes. In Fig. 12.15a, the intersection of the 2-cells A and B is the 1-cell depicted by thick lines. It is part of the boundary of both A and B. The intersection of the two 3-cells in Fig. 12.15b is given by the dark colored 2-cell and the 0 and 1-cells in its boundary. This structure is part of the boundary of both 3-cells. A counterexample is depicted in Fig. 12.16. Two solids penetrate, and hence, the intersection of both is not a common boundary; the structure is not a cell complex.

The advantages of representing GIS data as a cell complex are as follows.

- The model implies that there are no penetrations or overlaps of the interiors of cells; cells touch at least in common boundaries. This is an essential consistency constraint for many GIS applications; for example, two parcels do not overlap, and two buildings do not penetrate.
- The explicit representation of any touching between objects, i.e., the boundary and coboundary relations, facilitates navigational access to all neighboring objects, without the need to consider geometry. This supports the efficient processing of queries involving topological predicates, e.g., inside or touch. These predicates were discussed in the previous section.

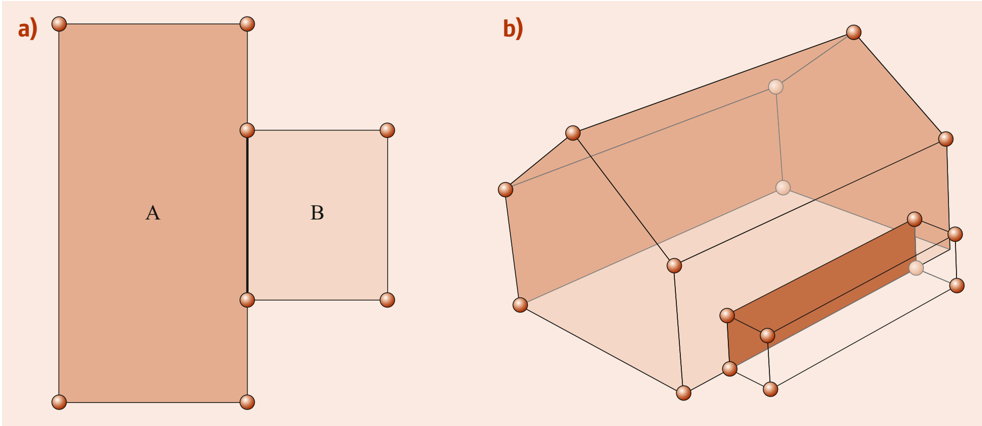


Fig. 12.15 Cell complexes, (**a**) consisting of two 2-cells A and B touching in a common edge (*depicted bold*), and (**b**) of two 3-cells touching in a common face (*depicted dark*)

All topological data models reviewed in the next sections are based on the concept of cell complexes. They differ with regard to the dimension of the embedding space (2-D or 3-D), the dimension of the cells, the geometric shape of the cells (triangles/tetrahedrons, arbitrary shape, enclaves), whether cells are explicitly or implicitly modeled, and to which degree the boundary and coboundary relations are modeled explicitly.

2-D Models

Realizations of 2-D topological models are the *maps* defined by *Plümer and Gröger* [21], which define a complete coverage of the plane by faces and are characterized axiomatically, i.e., provide an efficient and effective method to check whether a dataset has the map properties. In *Gröger* [22], the concept is extended by allowing holes in faces. *Molenaar's* [23, 24] *single* and multivalued vector maps are a special case of the 3-D-version, which will be described in the next section. The cells of the model presented by *Egenhofer* et al. [25] are restricted to triangles geometrically, whereas the *coverages* data type of Esri's GIS tools *ArcGIS* allow for faces of arbitrary shape, which may contain holes.

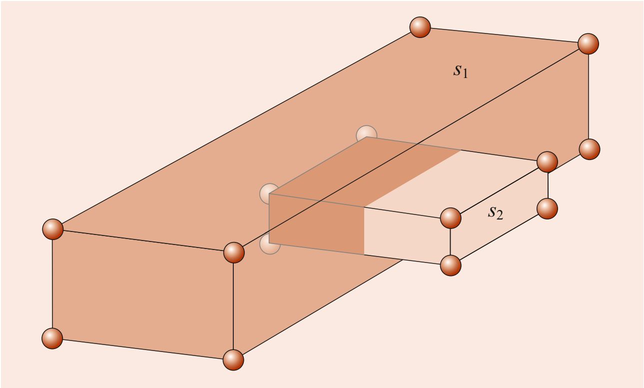


Fig. 12.16 0-cells, 1-cells, 2-cells, and 3-cells not constituting a cell complex; the intersection of solids s_1 and s_2 is not a cell in the boundary of both cells

Topological Networks

A *topological network* is a cell complex consisting of 0- and 1-cells, which is embedded in 3-D space. Another term for a topological network is a *graph* embedded in 3-D space. The focus of topological networks is on explicit modeling of the connectivity between line objects (edges) and junctions (nodes), not on surface topology. The main application area of topological networks is the modeling of transportation or utility networks (Chap. 26). The third dimension is often not represented explicitly, but due to overpasses and underpasses, 2-D or 2.5-D models are not sufficient. A prominent and widely used example is the standard *Geographic Data Files* (GDF) [26], which are the base of all data models for commercial vehicle navigation. Level 1 in GDF, which is used for path finding, is a topological network. Another example for a topological network is the graph representing reachability inside buildings, which is used for indoor path finding [27]. The representation of topological networks in data bases is discussed in Chap. 3.

3-D Models

A survey of 3-D topological models for GIS can be found in the work of *Zlatanova* et al. [28], whereas in the work of *Ellul* and *Haklay* [29], the requirements and benefits of such models for GIS applications are identified. The first topological data model in GIS from a historical perspective was the *formal data structure* (*FDS*) presented by *Molenaar* [30]. He distinguishes the primitive *nodes*, *arcs/edges*, and *faces*. Volumes are called *bodies* and exist on a feature level. Faces are bounded by edges/arcs, and each arc has a start and an end node. Each face has a body on the left-hand side and one on the right-hand side (Fig. 12.17). Edges are straight lines geometrically, and faces are planar and may contain holes. *Flick* [31] extends FDS by introducing *bodies* as topological primitives. The *urban data model* (*UDM*) developed by *Coors* [32] and the *simplified spatial schema* [19] modify this model by omitting the explicit representation of edges, thus facilitating efficient visualization. For the same reason, faces in UDM are restricted to triangles. A topological model

12

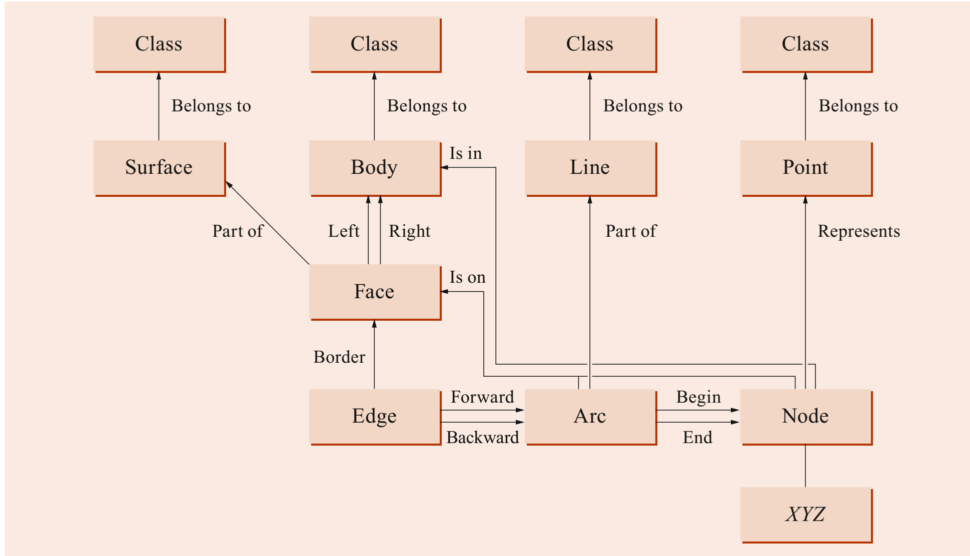


Fig. 12.17 Diagram of 3-D FDS by *Molenaar*. (Adapted from [23])

based on *simplicial complexes* – the restriction of cell complexes to triangles and tetrahedrons – is the *TEN* (tetrahedral network) structure [33]. It can be implemented in relational databases very efficiently [34].

The topological model introduced by *Pigot* [35] provides a full implementation of the concept of cell complexes, including all boundary and coboundary relations. The model defined by *Gröger* and *Plümer* [36] extends cell complexes in two respects: connectivity is considered as an additional requirement, prohibiting floating buildings, for example, and two special solids are introduced: a solid representing the air space and one representing the Earth's mass. Both are bounded only partially. Hence, this model defines a 3-D tessellation of space by solids: each point in 3-D space is in the

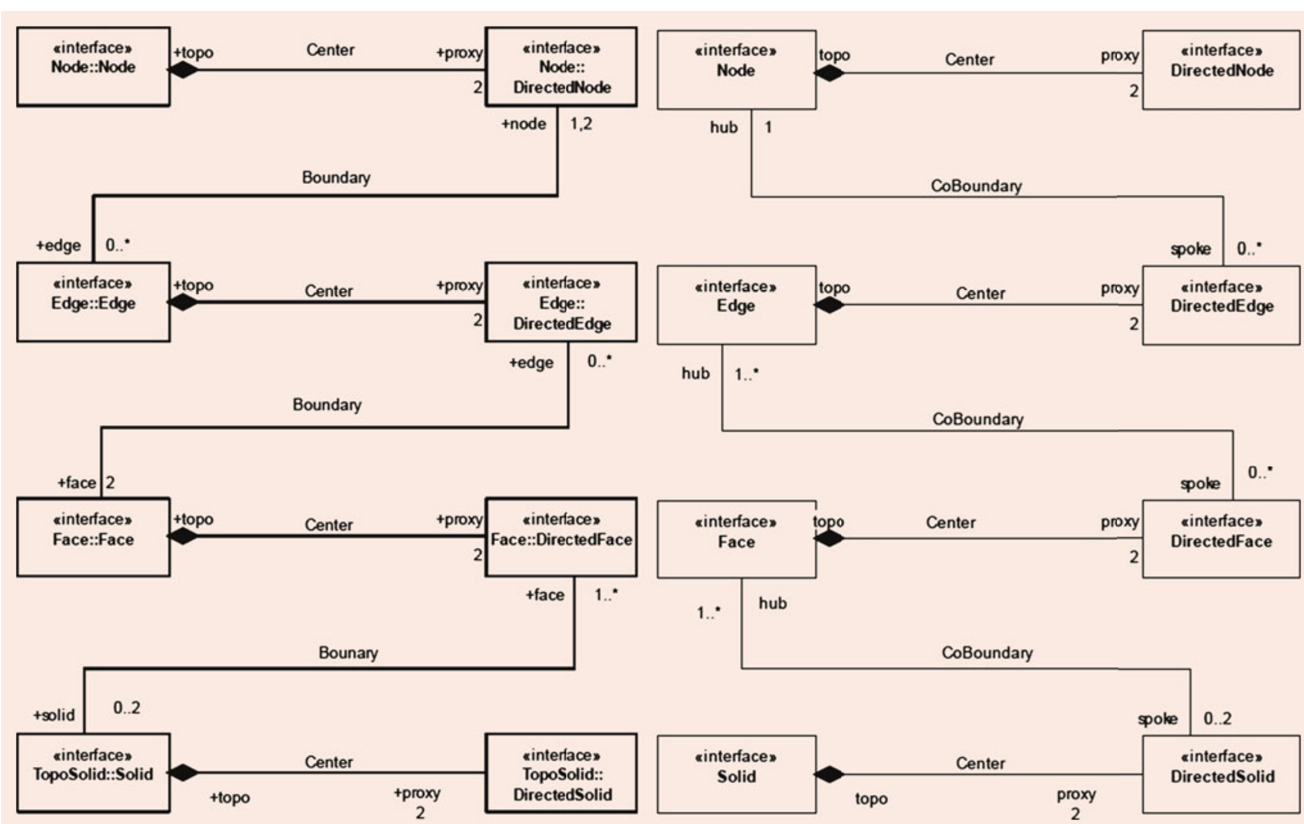


Fig. 12.18 Boundary (*left*) and coboundary (*right*) relations. (Adapted from [37])

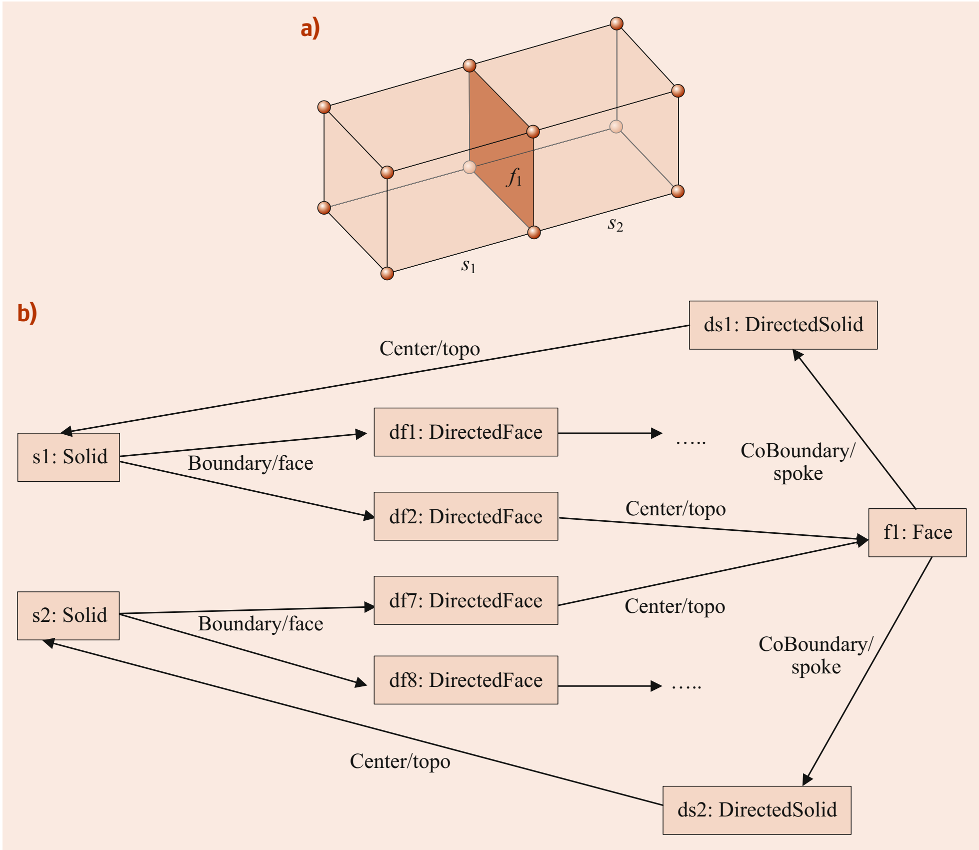


Fig. 12.19 **(a)** A 3-D scene with two solids s_1 and s_2 sharing a face f_1 and **(b)** an extract from the corresponding UML instance diagram

boundary of a solid or in the interior of exactly one solid. The declarative definition of the model is accompanied by *axioms*, which are used to check effectively and efficiently whether datasets are consistent, i.e., meet the requirements of the model. Transaction rules for updating datasets while preserving consistency are sketched in *Gröger and Plümer* [36].

A further topological model is provided by the standard ISO 19107 *Spatial schema* [4]. The model defines topological primitives for all dimensions (nodes, edges, faces, topological solids) and fully realizes the boundary and coboundary relations. The properties of the topological primitive are defined by their geometrical counterparts, which were described in Sect. 12.1; faces (interface *Face*) must be connected and may have holes delimited by interior rings, and topological solids (interface *Solid*) must also be connected and may have enclaves bounded by interior shells. All boundary and coboundary relations are represented explicitly (see the UML diagram in Fig. 12.18). In analogy to the orientable primitives on the geometry level, directed topology objects interconnecting the topological primitives are used to define consistently oriented boundaries and coboundaries (Fig. 12.18). For example, the boundary of a *TP Solid* consists of a set of directed faces (class *TP DirectedFace*); each directed face is assigned to exactly one *TP Face* by the *topo* role of the *center* association. This face is related to exactly one other directed face, which represents the face's role in the boundary of another topological solid that is a neighbor of the first one. Vice versa, the coboundary relations are defined by using directed topology objects; a face, for example, has a coboundary relation to two directed solids, each of them relating to a solid that is bounded by the face (Fig. 12.19).

In addition to the boundary and coboundary relations, where the difference of dimensions of the cells is 1, there is a relation called *isolated*, which associates a node (0-cell) with a face (2-cell) or a topological solid (3-cell), when this node is inside the interior of the face or of the topological solid. Likewise, an edge is related to a topological solid by that association, when the edge is completely in the interior of the solid.

ISO 19107 provides classes for defining topologies (prefix *TP*) that are independent of its geometrical counterparts (prefix *GM*) but are related by associations. The advantage of this approach is flexibility; there are three options to use topology.

1. Topology is omitted, i.e., only the geometrical aspects are represented.
2. Both geometry and topology are modeled and linked, combining the advantages of both representations.
3. Only topology is represented, i.e., a scene is represented purely structural by topological primitives and its boundary and coboundary relations, without any (geo)metrical information like shape, size, or location.

If the purely topological representation in case 3 is restricted to point and line primitives and the corresponding boundary and coboundary relations, one obtains the well-known *graph* structure. This structure and the corresponding algorithms, which are crucial for GIS, are the topic of Sect. 12.3.

12.3 Graph Theory and the Königsberg Bridge Problem

In geographic information systems, concepts from graph theory are extremely useful in expressing the spatial structure of entities seen as points, lines, areas and solids, after the geometrical details of these entities are removed. For example, in transportation and river networks, the topological properties of their structures can be represented using graphs. This article describes the origins of graph theory and the impact it has on various fields ranging from geography to economics.

The Königsberg bridge problem is a classic problem based on the topography of the city of Königsberg, formerly in Germany but now known as Kaliningrad and part of Russia. The river Pregel divides the city into two islands and two banks as shown in Fig. 12.20.

The city had seven bridges connecting the mainland and the islands (represented by thick lines in the Fig. 12.20; [38–41]). The problem is whether there is a walk that starts at any island, traverses every bridge exactly once, and returns to the starting point. The solution proposed by a Swiss mathematician, Leonhard Euler, led to the birth of a branch of mathematics called graph theory, which finds applications in areas ranging from engineering to the social sciences. Euler proved that there is no solution to the problem based on the number of bridges connecting each land area.

The results from the solution of the Königsberg problem have been extended to various concepts in graph theory. In graph theory, a path that starts and ends at the same node and traverses every edge exactly once is called a Eulerian circuit. The result obtained in the Königsberg bridge problem has been generalized as Euler's theorem, which states that a graph has a Eulerian circuit if and only if there are no nodes of odd degree. Since the graph corresponding to Königsberg has four nodes of odd degree, it cannot have a Eulerian circuit. Subsequently, the concept of Eulerian paths was introduced, which deals with paths that traverse every edge exactly once. It was proved that such a path exists in a graph if and only if the number of nodes of odd degree is two [40–44]. A node is a "node of odd degree" if the number of edges incident to the node is odd.

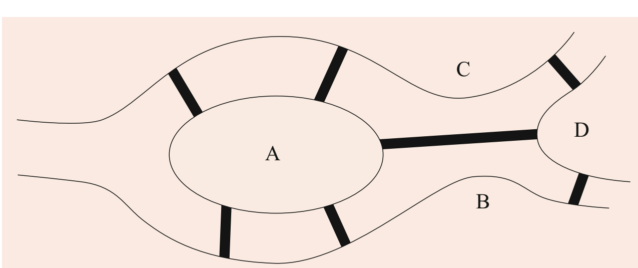


Fig. 12.20 Layout of the city of Königsberg showing the river, bridges, and land areas A, B, C, and D

While studying the Königsberg bridge problem, Euler also observed that the number of bridges at every land area would add up to twice the number of bridges. This result came to be known as the hand-shaking lemma in graph theory, which states that the sum of node degrees in a graph is equal to twice the number of edges. This result is the first formulation of a frequently used result in graph theory that states that the sum of node degrees in a graph is always even [43, 44].

12.3.1 Abstraction

The Königsberg bridge problem was formulated based on the layout of the city of Königsberg around the river Pregel. The problem was to find a tour that starts at any point in the city, crosses each bridge exactly once, and returns to the starting point. No one succeeded in doing this.

Euler formulated the problem as finding a sequence of letters A, B, C, D (that represent the land areas), such that the pairs (A, B) and (A, C) appear twice (thus representing the two bridges between A and B, and A and C) and the pairs (A, D), (B, D), (C, D) appear only once (these pairs would represent the bridges between A and D, B and D, C and D). Euler used a counting argument to prove that no such sequence exists, thus proving that there the Königsberg bridge problem has no solution. Euler presented this result in the paper, "The solution of a problem relating to the geometry of position" at the Academy of Sciences of St. Petersburg in 1735. This paper, in addition to proving the nonexistence of a solution to the Königsberg bridge problem, gave some general insights into arrangements of bridges and land areas [42, 43, 45].

Euler summarized his main conclusions as follows.

1. If there is any land area that is connected by an odd number of bridges, then a cyclic journey that crosses each bridge exactly once is impossible.
2. If the number of bridges is odd for exactly two land areas, then there is a journey that crosses each bridge exactly once is possible, if it starts at one of these areas and ends in the other.
3. If there are no land areas that are connected by an odd number of bridges, the journey can start and end at any land area [43].

Euler gave heuristic reasons for the correctness of the first conclusion. To complete a cyclic journey around the land areas, crossing each bridge exactly once, there must be a bridge

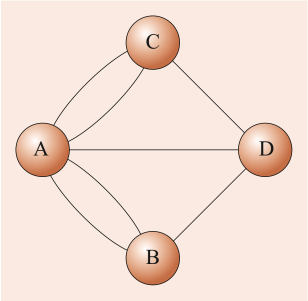


Fig. 12.21 Graph representation of the city of Königsberg

to leave the area for every bridge to enter it. This argument was generalized to the conclusion that a cyclic journey is possible if every island is connected by an even number of bridges. Formal proofs for the conclusions were not proposed until the year 1871, in a posthumous paper by Carl Hierholzer [39, 42].

The paper presented by Euler on the Königsberg bridge problem can be considered to mark the birth of graph theory, in general. Later, a diagrammatic representation evolved, which involved nodes or vertices, and the connecting lines that are called edges. Using this representation, the Königsberg problem is modeled as shown in Fig. 12.21.

Circles, called nodes, represent the islands and the banks, and connecting lines called edges represent the bridges. The number of edges that are incident on a node is called the degree of the node [43].

In the Königsberg bridge problem, the number of bridges connecting a land area would be the degree of the node representing the land area.

In an undirected graph, a cycle that traverses every edge exactly once is called a Euler tour or Euler cycle. Any graph that possesses a Euler cycle is called a Eulerian graph. A path that traverses each edge exactly once with different starting and end points is called a Eulerian path. An undirected multigraph has a Eulerian circuit (path) if and only if it is connected, and the number of vertices with odd degree is zero (two).

Figure 12.22 illustrates the Eulerian path and the Eulerian cycle in a graph. In Fig. 12.22a, a Eulerian path exists, and it can be observed that the graph has exactly two, odd degree vertices, which would be the start and end vertices of the Eulerian path, A–B–C–D–A–C. Figure 12.22b does not have vertices with odd degree and has a Eulerian cycle, whereas Fig. 12.22c has neither a Eulerian path nor a Eulerian cycle.

12.3.2 Finding a Eulerian Circuit in a Graph

The method successively finds cycles in the graph. At each step, the edges that are in the already discovered cycles are removed, and the cycle is spliced with the one discovered in the previous step. This process is continued until all edges are exhausted. These basic ideas were formalized into an algorithm in [46]. The algorithm maintains a list L with each vertex x, such that the k-th entry in the list indicates the vertex to visit when vertex x is reached the k-th time.

Algorithm

Step 1: Select any vertex $v1$. $v = v1$; $k[v] = 0$ for all vertices v; set $kv = 0$. Label all edges as unvisited.

Step 2: Select an unvisited edge e incident to v. Mark this edge "visited". Let w be the other end vertex of e. Increment kv by 1, and $Lv[kv] = w$. If w has an unvisited incident edge, go to step 3. If not, y will be $v1$. Then, go to Step 4.

Step 3: Set $v = w$ and go to Step 2.

Step 4: Find a vertex $v1$ such that there is at least one visited edge and one unvisited edge incident at $v1$. Set $v = v1$ and go to Step 2. If no such vertex exists, go to Step 5.

Step 5: To construct the Eulerian circuit, start at $v1$. The first time a vertex u is reached, proceed to the vertex $Lu[ku]$. Decrement ku and continue.

Trace of the algorithm for Fig. 12.23

Step 1: $v1 = 1 = v$; $kx = 0$ for $x = 1, 2, 3, 4$.
Step 2: Select edge a. $w = 2$; $k2 = 1$; visited$(a) = 1$.
Step 3: $v = 2$; Select edge b. $w = 3$; $k3 = 1$; visited$(b) = 1$.
Step 4: $v = 3$; Select edge c. $w = 4$; $k4 = 1$; visited$(c) = 1$.
Step 5: $v = 4$; Select edge d. $w = 1$; $k1 = 1$; visited$(d) = 1$.
Step 6: $v = 2$;
Step 7: Select edge e; $w = 4$; $k4 = 2$; visited$(e) = 1$
Step 8: $v = 4$;
Step 9: Select edge f; $w = 2$; $k2 = 2$; visited$(f) = 1$
Step 10: Construct the cycle as 1–2–4–2–3–4–1.

12.3.3 Famous Applications

Eulerian cycles find applications in problems where paths or cycles need to be found that traverse a set of edges in a graph. Such problems are generally called edge routing problems.

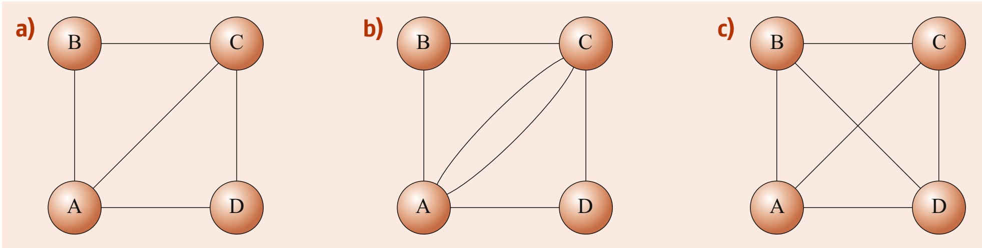


Fig. 12.22 Illustration of a Eulerian path and a Eulerian cycle. **a** Eulerian path: A–B–C–D–A–C, **b** Eulerian cycle: A–B–C–D–A–C–A, **c** neither Eulerian path nor cycle exist

12

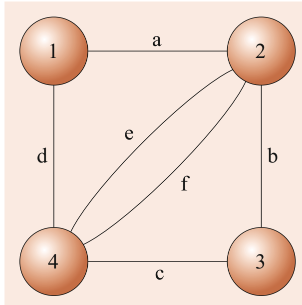


Fig. 12.23 The Eulerian algorithm

Snow Plow Problem

This problem requires finding the least distance route in the road network that starts and ends at the station, so that snow can be cleared from the streets at minimum cost. The minimum distance route is obviously the Eulerian cycle, since this cycle traverses each edge exactly once. However, it is unlikely that any real road network would happen to satisfy the necessary conditions that make it Eulerian. In that case, the problem moves to the realm of "the Chinese postman problem" [46–48].

Chinese Postman Problem

A postman delivers mail everyday in a network of streets. It is useful to know whether or not the postman can traverse the network and return to the mail station without driving the length of any street more than once. If the network is not Eulerian, the problem gets modified to the one where it is required to find the shortest path, which visits each edge at least once. This problem statement requires a parameter to be associated with each edge that represents the cost of traversing that edge. For example, cost can be represented in terms of the length of the street, which the edge represents.

In a non-Eulerian graph, the postman's circuit, the shortest or otherwise, will repeat one or more edges. Every vertex is entered the same number of times that it is left, so that any vertex of odd degree has at least one incident edge that is traversed at least twice. The Chinese postman problem is formulated as an optimization problem, where the total cost of repeated edges is minimized.

Algorithm

Step 1: Find the shortest path between each pair of odd degree.

Step 2: Find the subgraph G′ with odd degree vertices.

Step 3: Find the minimum weight matching of all the edges in G′. The edges in the shortest path connecting a matched pair of odd degree vertices should be repeated.

Figure 12.24 shows a sample graph with edge weights, and the Chinese postman algorithm finds the least cost (minimum edge weight) path in the graph, such that every edge is traversed at least once. Table 12.1 shows the shortest path

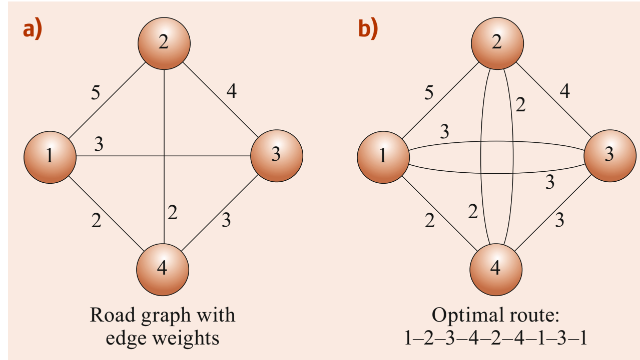


Fig. 12.24 The Chinese postman problem algorithm. **a** Road graph with edge weights, **b** optimal route: 1–2–3–4–2–4–1–3–1

Table 12.1 The shortest path cost between the pairs

	1	**2**	**3**	**4**
1	0	5	3	2
2	5	0	4	2
3	3	4	0	3
4	2	2	3	0

Table 12.2 Matchings and costs

Matching	Cost
(1, 2), (3, 4)	$5 + 3 = 8$
(1, 4), (2, 3)	$2 + 4 = 6$
(1, 3), (2, 4)	$3 + 2 = 5$

costs between every pair of vertices, which is used by the algorithm to find the minimum weight matchings on edges (the matchings and their costs are listed in Table 12.2). The algorithm finds that the paths from vertex 1 to vertex 3, and the path from 2 to 4 must be repeated, since this is the minimum cost matching (the cost is 5). The algorithm finds the optimal route to be 1–2–3–4–2–4–1–3–1 in the graph shown in Fig. 12.24.

Capacitated Chinese Postman Problem

This problem arises where each edge has a demand, and vehicles to be routed have finite capacities. For example, in applications involving road salting in the winter season, there is a limit on the maximum amount of salt that a truck can carry. The amount of salt required is fixed for a road segment. The capacitated Chinese postman problem finds the sets of routes from a single station that service all the road segments in the network at a minimal cost and subject to the constraint that the total demand of each route does not exceed the capacity of each truck. *Christofides* [47] proposed an algorithm to solve this problem.

Capacitated Arc Routing Problem

This problem is different from the capacitated Chinese postman problem in that demands of some of the road segments can be zero. This situation can arise in road salting scenarios where state highways can be used for traveling but need not be salted. These edges can be used to traverse between the edges that require the service.

Both the capacitated Chinese postman problem and capacitated arc routing problem are NP-hard [48], and heuristic methods are normally used to obtain solutions.

12.3.4 Graph Theory

The Königsberg problem had a powerful impact on mathematics, paving the way for the creation of a new modeling theory called graph theory. The applications of graph theory are numerous in science and engineering. A few are listed below.

Graph Theory in Spatial Networks

The very fact that graph theory was born when Euler solved a problem based on the bridge network of the city of Königsberg points to the apparent connection between spatial networks (e.g., transportation networks) and graphs. In modeling spatial networks, in addition to nodes and edges, the edges are usually qualified by adding weights that encode information like the length of the road segment that the edge represents. Connectivity and shortest paths in spatial networks have been extensively studied using graphs [49].

Spatiotemporal networks: the addition of a temporal dimension to a spatial network certainly enhances its analysis capabilities, in light of the fact that most networks exhibit a time dependence. For example, the travel times on road segments display variations with time of day. Incorporating the temporal variations into network analysis would definitely take the model further in its effectiveness in addressing user queries. For example, a user would be able to ask questions such as "Which route would take me fastest to my work, if I start from home at 8 am?" or "When is the best time to go to the grocery store so that I spend the least time driving?". Considering the time-dependence of most applications modeled as networks, accounting for the temporal factor would help answer queries more realistically. Adding time to the analysis would also help define time windows for operations such as visiting a list of nodes within given time intervals at a minimum cost, such as in the case of the traveling salesman problem [50].

Shortest Path

A problem that is frequently encountered in the context of spatial networks is that of finding the shortest path from one location to another. In a spatial network that is modeled as a graph, this amounts to finding the shortest path from a node (the start node) to another (the end node), and the path would be represented as an ordered sequence of nodes and edges starting from the start node to the end node [51, 52]. In cases where the spatial networks display time dependence, this temporal aspect would need to be accounted for in shortest-path computations. For example, in a transportation network, the shortest path could possibly depend on the start time [50].

Traveling Salesman Problem (TSP)

A computation that has widespread applications is finding an optimal route that touches multiple locations in the network, which is known as the traveling salesman problem (TSP). Although finding an optimal solution to TSP is NP-complete, there are heuristic approaches that are computationally efficient and guarantee good quality solutions. Two variations of simple TSP are closed TSP, where the start and end nodes of the solution are the same, and open TSP, where the start node is not the same as the end node. A significant number of real-life applications might require enhancements to accommodate factors such as restrictions on time windows during which nodes are visited and the order of visits.

Vehicle Routing Problem (VRP)

An optimization problem called the vehicle routing problem (commonly known as VRP) has gained significant attention since it was formulated in 1950s [53]. The interest in this problem has expanded from one of research interest to practical applications such as home delivery services and postal services. In its simplest form, VRP is an optimization problem that finds optimal routes for m vehicles to deliver discrete quantities of goods to n customers. The goods are assumed to be stored in a depot, which is also the initial location of the vehicles (Fig. 12.25). Optimization, in most cases, focuses on minimizing the total distance traveled by vehicles (or the total driving time). In some cases, the number of vehicles is also considered to be an optimization variable.

Real-world problems bring more dimensions to the problem, and VRP has been modified to accommodate the re-

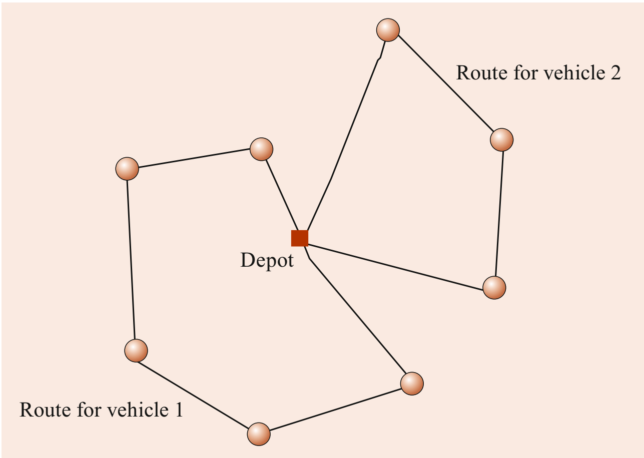


Fig. 12.25 Simple example that illustrates a VRP solution. It allocates a group of customers to each of the two vehicles and finds the optimal route that visits each customer

12

quirements of application scenarios. The modified versions of VRP, often termed rich VRP formulations, include cases where there are constraints on capacities of vehicles (capacitated vehicle routing problem, CVRP) and time windows in which the customers can be served (vehicle routing problem with time windows, VRPTW). In addition to these, we do see extended formulations of VRP, such as multiple depot VRP, which brings in a new dimension of inventory and periodic VRP. Though the optimal solution to these problems is NP-hard, heuristic solutions that exploit the properties of application domains that guarantee good quality solutions are available [54].

Since the optimal routes are typically computed on a transportation network that is modeled as a graph, this problem falls well within the purview of spatial networks and graphs. One popular approach to solving VRP on a road network is to initially cluster the customers based on spatial proximity and compute a route that visits all customers, often using traveling salesman problem solutions (Fig. 12.25).

Graphs in Sociology

Social structures represented as graphs are generally known as social networks. This helps in investigating and characterizing social structures using graph theory techniques. Individual actors, mostly persons, are modeled as nodes, and the relationships and interactions among them are represented as links. Social networks are widely used to model structures such as friendship and acquaintance networks and business networks. In addition to purely sociological interest, they have extended to areas such as epidemiology, where they employ social network tools to study disease transmission, economics, political science, and sociolinguistics.

Important social network metrics include betweenness, centrality, degree, and page rank. Algorithms to compute these measures rely heavily on the graph structure that models the social structures.

Graph Theory in Geography

Graphs are also widely applied in geography in the modeling of stream systems. Streams have been modeled as hierarchical graphs and random graphs in the literature [55].

In addition to the applications described above, graphs find other wide applications, including modeling of social networks, molecular structures in chemistry, computer networks, electrical networks, and syntax structures in linguistics.

12.3.5 Future Directions

Relationships between Eulerian graphs and other graph properties, such as the Hamiltonian property, are being studied [56]. The graph, i.e., the mathematical model that owes its origin to the Königsberg bridge problem, is being extended from its static form to a dynamic representation due to its applications in domains where structures evolve (such as in social networks) or in cases where graph properties change (an example being a transportation network).

References

1. Foley, J.D., van Dam, A., Feiner, S.K., Hughes, J.F.: Computer Graphics: Principles and Practice, 2nd edn. Addison Wesley, Boston (1995)
2. Mäntylä, M.: An Introduction to Solid Modeling. Principles of Computer Science. Computer Science, New York (1988)
3. Mortenson, M.E.: Geometric Modelling. Wiley, Chichester (1997)
4. Herring, J.: The OpenGIS Abstract Specification, Topic 1: Feature Geometry (ISO 19107 Spatial schema). OGC Document Number 01-101. (2001)
5. Cox, S., Cuthbert, A., Lake, R., Martell, R.: OpenGIS Geography Markup Language (GML). Open GIS Doc. No. 02-009. (2002)
6. Lake, R., Burggraf, D., Trninic, M., Rae, L.: Geography Mark-Up Language: Foundation for the Geo Web. Wiley, Chichester (2004)
7. Portele, C.: OpenGIS Geography Markup Language (GML) encoding standard, v.3.2.1. OGC Doc. No. 07-036. (2007)
8. Armstrong, M.A.: Basic Topology. Springer, New York (1997)
9. Bungartz, H.-J., Griebel, M., Zenger, C.: Einführung in die Computergraphik. Grundlagen, Geometrische Modellierung, Algorithmen. Vieweg, Wiesbaden (2002). in German
10. Gröger, G., Kolbe, T.H., Nagel, C., Häfele, K.H.: OpenGIS City Geography Markup Language (CityGML) Encoding Standard, v 2.0.0. OGC Doc. No. 12-019. (2012)
11. Vretanos, P.A.: OpenGIS Filter Encoding Implementation Specification (2005)
12. Oracle Cooperation: Oracle Spatial Developers Guide 11g release 1 (11.1) (2009)
13. Alexandroff, P.: Elementary Concepts of Topology. Dover, New York (1961)
14. Hatcher, A.: Algebraic Topology, 10th edn. Cambridge Univ. Press, Cambridge (2001)
15. Cohn, A.G., Bennet, B., Gooday, J., Gotts, M.M.: Qualitative spatial representation and reasoning with the region connection calculus. Geoinformatica **1**, 275–316 (1997)
16. Egenhofer, M.J., Franzosa, R.D.: Point-set topological spatial relations. Int. J. Geogr. Inf. Sci. **5**(2), 161–174 (1991)
17. Egenhofer, M.J., Herring, J.R.: Categorizing Binary Topological Relationships Between Regions, Lines and Points in Geographic Databases. Technical Report. Department of Surveying Engineering, University of Maine, Orono (1991)
18. Clementini, E., Di Felice, P.: A comparison of methods for representing topological relationships. Inform. Sci. **3**(3), 149–178 (1995)
19. Zlatanova, S.: 3D GIS for Urban Development. International Institute for Aerospace Survey and Earth Sciences; Institute for Computer Graphics and Vision Graz University of Technology, Enschede, Graz (2000)
20. Herring, J.R.: OpenGIS Implementation Specification for Geographic information – Simple Feature Access – Part 1: Common architecture: v.1.2.0. OGC Doc. No OGC 06-103r3. (2006)
21. Plümer, L., Gröger, G.: Achieving integrity in geographic information systems-maps and nested maps. Geoinformatica **1**(4), 345–367 (1997)
22. Gröger, G.: Modellierung raumbezogener Objekte und Datenintegrität in GIS. Wichmann, Heidelberg (2000). in German

23. Molenaar, M.: Formal data structures, object dynamics and consistency rules. In: Ebner, C., Fritsch, D., Heipke, C. (eds.) Digital Photogrammetric Systems. Wichmann, Heidelberg (1991)
24. Molenaar, M.: An Introduction to the Theory of Spatial Object Modelling for GIS. Taylor Francis, New York (1998)
25. Egenhofer, M.J., Frank, A.U., Jackson, J.P.: A topological data model for spatial databases. In: Des. Implement. Large Spat. Databases, First Symp. SSD '89 Santa Barbara, 1989. Springer, Berlin, Heidelberg (1990). Proceedings
26. ISO: ISO 14825:2004 Intelligent Transport Systems – Geographic Data Files (GDF) – Overall Data Specification. ISO, Geneva (2004)
27. Becker, T., Nagel, C., Kolbe, T.H.: A multilayered space-event model for navigation in indoor spaces. In: Lee, J., Zlatanova, S. (eds.) 3D Geo-Information Sciences, pp. 61–78. Springer, Berlin, Heidelberg (2009)
28. Zlatanova, S., Abdul Rahman, A., Shi, W.: Topological models and frameworks for 3D spatial objects. Comput. Geosci. **30**, 419–428 (2004)
29. Ellul, C., Haklay, M.: Requirements for topology in 3D GIS. Trans. GIS **10**(2), 157–175 (2004)
30. Molenaar, M.: A topology for 3D vector maps. ITC J. **1**, 25–33 (1992)
31. Flick, S.: Konzeption eines adaptiven Frameworks für 3D-Geo-Informationssysteme. Fraunhofer IRB, Stuttgart (1999). in German
32. Coors, V.: 3D-GIS in networking environments. Comput. Environ. Urban. Syst. **27**, 45–357 (2003)
33. Pilouk, M.: Integrated Modelling for 3D GIS. ITC Publication, vol. 40. ITC, Enschede (1994). Dissertation
34. Penninga, F., Oosterom, P.: A compact topological DBMS data structure for 3D topography. In: Fabrikant, S.I., Wachowicz, M. (eds.) The European Information Society: Leading the Way with Geo-information, pp. 455–471. Springer, Berlin, Heidelberg (2007)
35. Pigot, S.: A topological model for a 3D spatial information system. In: Proc. 5th Int. Symp. Spat. Data Handl. Charleston. pp. 344–360. (1992)
36. Gröger, G., Plümer, L.: Provably correct and complete transaction rules for updating 3D city models. Geoinf. **17**(1), 131–164 (2012)
37. ISO: ISO 19107:2019 Geographic Information – Spatial Schema, 2nd edn. ISO, Geneva (2019)
38. Chartrand, G.: The Königsberg Bridge problem: an introduction to Eulerian graphs. In: Introductory Graph Theory. Dover, New York (1985)
39. Harary, F.: Graph Theory. Addison-Wesley, Reading, pp. 1–2 (1994)
40. Biggs, N.L., Lloyd, E.K., Wilson, R.J.: Graph Theory 1736–1936. Oxford Univ. Press, Oxford (1976)
41. Bollobás, B.: Graph Theory: An Introductory Course. Springer, Berlin, Heidelberg, p. 12 (1979)
42. Steinhaus, H.: Mathematical Snapshots, 3rd edn. Dover, New York, pp. 256–259 (1999)
43. Biggs, N.L., Lloyd, E.K., Wilson, R.J.: Graph Theory. Oxford Univ. Press, Oxford (1998)
44. Wilson, R.J.: Introduction to Graph Theory. Longman, New York (1985)
45. Hartsfield, R.: Pearls in Graph Theory. Academic Press, San Diego (1990)
46. Edmonds, J., Johnson, E.L.: Matching, Euler tours and Chinese Postman. Math. Program. **5**, 88–124 (1973)
47. Christofides, N.: The optimum traversal of a graph. Int. J. Manag. Sci. **1**(6), 719–732 (1973)
48. Evans, J.R., Minieka, E.: Optimization Algorithms for Networks and Graphs. Marcel Dekker, New York (1992)
49. Shekhar, S., Chawla, S.: Spatial Databases: A Tour. Prentice Hall, Upper Saddle River (2003)
50. George, B., Kim, S., Shekhar, S.: Spatio-temporal Network Databases and Routing Algorithms: A Summary of Results. In: Proc. 10th Int. Symp. Spat. Temporal Databases (SSTD) 07.2007. (2007)
51. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numer. Math. **1**, 269–271 (1959)
52. Hart, P.E., Nilsson, N.J., Raphael, B.: A formal basis for the heuristic determination of minimum cost paths. IEEE Trans. Syst. Sci. Cybern. **4**(2), 100–107 (1968)
53. Dantzig, G., Ramser, J.: The truck dispatching problem. Manag. Sci. **6**, 80–91 (1959)
54. Golden, B.L., Raghavan, S., Wasil, E.A.: The Vehicle Routing Problem: Latest Advances and New Challenges. Springer, Heidelberg (2008)
55. Wilson, R.J., Beineke, L.W. (eds.): Applications of Graph Theory. Academic Press, New York (1979)
56. Fleischner, H.: (Some of) the many uses of Eulerian Graphs in graph theory. Discrete. Math. **230**, 23–43 (2001)

Gerhard Gröger Gerhard Gröger received a Diploma in Computer Science from the University of Bonn, a PhD degree in Environmental Sciences from the University of Vechta, and the facultas docendi/venia legendi for Geoinformation from the University of Bonn. Currently he is a project leader and software designer at CPA ReDev GmbH, Siegburg, Germany. He gives lectures at the University of Bonn (BSc/MSc Geodesy and Geoinformation).

Betsy George Betsy George currently works in the Spatial Databases Group at Oracle Corporation after earning her PhD from Department of Computer Science at University of Minnesota, Twin Cities, USA. Her research interests include spatial databases with specific focus on spatiotemporal networks and routing algorithms in time-dependent networks.

Cartography

13

Aileen Buckley, Paul Hardy, and Kenneth Field

Contents

A. Buckley (✉)
Esri
Redlands, CA, USA
e-mail: abuckley@esri.com

P. Hardy
Cambridge, UK
e-mail: paul@paulhardy.net

K. Field
Esri
Redlands, CA, USA
e-mail: kfield@esri.com

Abstract

This chapter reviews the geographic data that underpins a map, the process required to portray that information in graphic form, basic principles of map design, and considerations for page layout. While this is not a comprehensive review of all aspects of cartography, several excellent resources that you can refer to later are cited throughout the chapter and in the References. After reading this chapter, you should find that you are better able to understand what a map shows and how it is shown. You should also be more critically aware of how information can be portrayed effectively on maps. This chapter will also help you better appreciate cartography as a crucial part of a geographic information system (GIS).

A review of the development of cartography sets the scene for the current state of the art and science in this field (Sect. 13.1). The principles and practices of cartography are presented to provide a clear rationale for the importance of map design in communicating relevant geographic information. Looking at the various types of maps and their construction allows you to consider the constraints on map design and the consequences for how data is portrayed on maps (Sect. 13.2). A review of the design and use of symbols, color, and type on maps illustrates the many possibil-

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_13

ities and choices available to the mapmaker (Sects. 13.2–13.6). Relief portrayal techniques illustrate how data about a terrain surface can be used to create a three-dimensional representation of the physical environment (Sect. 13.7). The final sections cover the intricacies of map design and page layout (Sects. 13.8 and 13.9). Map production and reproduction are not covered here.

13.1 Cartography in Review

Cartography is often defined as the study and practice of making maps. For this chapter, we concentrate on the practical aspects of mapmaking, which involve a combination of art, craft, and science. Cartography has continually evolved to use new technologies, of which GIS is among the most recent. Manual techniques that were the mainstay of cartographic design and production as recently as the 1980s have now been completely replaced by digital workflows, and new methods that were previously unavailable to or unimagined by cartographers are now pushing the frontiers of cartography.

This technological transformation has had consequences for cartography. Early innovations focused on providing high-quality design required for the full range of maps. GIS software interfaces and parameter defaults tended to lead the innocent user to produce garish and cluttered maps. Today, GIS routinely incorporates the accumulated best practices from cartographic practice and research and even pushes the limits of map design. As technology further evolves from the isolated desktop computer to server, mobile, and cloud environments, cartographic methods and practices have adapted yet again. However, the fundamental tenets of good map design remain largely unchanged; technology simply leads to different modes of implementation, as well as exciting new avenues in the creation of map-based products.

13.1.1 Brief History of Cartography

Cartography is one of the oldest of human graphic portrayal methods, dating back tens of thousands of years to Stone Age rock carvings showing routes to important geographic sites, such as good hunting grounds and foraging sites. Cartography progressed to include Babylonian clay tablet village plans; Polynesian stick charts for navigation (Chap. 14); Egyptian, Greek, and Chinese papyrus and parchment maps; medieval *mappa mundi* (on cloth); and portolan navigational charts on vellum. With the introduction of the printing press in the fifteenth century came paper. Computers have more recently been used to produce not only printed maps, but also ephemeral maps displayed on the screens of computers and mobile devices. Along the way, a wealth of knowledge has been accumulated about the best practices for making maps of different types. This chapter summarizes many of these practices, especially with regard to GIS and mapping.

13.1.2 Cartography and GIS

In its broadest sense, the term *cartography* can be described as the entire mapmaking process—from landscape survey, data collection, and information distillation to graphic compilation of the map and, finally, production and reproduction. In the context of GIS—a technology predicated on spatial data—the term *cartography* relates primarily to the graphic and production aspects of mapping. The early stages of data compilation and analysis are considered largely noncartographic in nature, although visualization of the data leads to initial insights by revealing unknowns [1]. The current view is that geographic analysis is about finding an answer to a geographic question, and cartography is about communicating that answer, generally to a larger audience than the analyst or the mapmaker. In this context, geographic analysis and cartography are complementary, and GIS is essential to both.

The relationship between GIS and cartography has not always been so harmonious. In the early days of GIS, excitement over the nascent technology led software developers to create computer mapping tools without investing time to learn from trained and experienced cartographers, who subsequently recognized and lamented the poor quality of GIS-designed maps. Users accepted the maps created with GIS software as being not only of sufficient quality but superior to traditional maps, simply because they were created with an innovation. The result was a breed of geospatial professionals who were more than capable of developing or using GIS but possessed little cartographic knowledge or training. However, in recent years, GIS has contributed to a rebirth in cartography, and cartography has helped elevate the stature of GIS. GIS tools are now capable of meeting the high-quality standards set by cartographers. In this Information Age, people have a greater appetite for information about the world. Thus map design and production have become relevant and vital not only for geospatial professionals but also for their map-reading audiences. People demand high-quality maps, and GIS developers and users have accepted the challenge and responsibility to ensure that the maps created with GIS technology meet these demands.

A high-quality map is much more than a visual report of GIS data—a well-designed map has elegance and style, clarity and meaning, and the appropriate content to convey the necessary level of precision and accuracy for the intended message. The responsibility of the cartographer is to ensure that the map communicates clearly, without clutter or confusion. A good mapmaker will often adhere to the axiom “Perfection is achieved not when there is nothing more to add, but when there is nothing left to take away.” (Antoine de Saint-Exupéry). A good example of a well-made map is the iconic London Underground Tube map by Harry Beck, which removes all aboveground detail other than the River Thames; dispenses with positional accuracy and scale uniformity in order to clearly show connectivity; and demonstrates an innovative design, use of color, and layout. It has thus become a cartographic classic.

Cartographic design is rarely right or wrong in an absolute sense, although some decisions clearly result in incorrect representations of information. Instead, map design can be viewed as more—or less—effective. The challenge for novice cartographers is to learn what constitutes effective cartographic design so that they can ensure that theirs are better maps.

13.2 Types of Maps

Maps take different forms and support a range of uses. Each map type has specific requirements in terms of design and compilation to communicate its message purposefully and support its proper use. It is difficult to categorize all maps neatly by a specific type; nonetheless, cartographers often classify maps into one of three broad categories: reference maps, charts, and thematic maps.

13.2.1 Reference Maps

A reference map (sometimes called a general-purpose map) displays both natural and cultural features in the geographic environment. These features are used to identify the spatial location of objects in an absolute (for example, in latitude and longitude coordinates) or relative sense and in relation to one another. Reference maps focus on the location of a variety of geographic features in an area, with limited focus on feature attributes (aside from type and, sometimes, name). Such a map provides a picture of the geographic character of the mapped area and the spatial configuration of features within that area (Fig. 13.1).

A globe is an example of a reference map that gives a general overview of the geographic nature of the entire earth. Reference maps can also give more detailed views of a portion of the earth’s surface. For example, reference maps found in atlases illustrate whole countries or continents, while the maps in a national map sheet series show smaller areas in greater detail.

A topographic map—so called because it portrays the topography or shape and features on the surface of the

Fig. 13.1 Portion of a reference map showing physical features of North America (courtesy of Tom Patterson, http://www.shadedrelief.com/north-america/)

Fig. 13.2 Portion of a 1 : 24 000-scale topographic map of San Francisco, California (courtesy of United States Geological Survey (USGS))

earth—is another example of a reference map (Fig. 13.2). Not only does it depict landforms, but it also includes other natural features, such as coastlines, rivers and creeks, vegetated areas, and lakes, along with cultural features, such as buildings, roads, cemeteries, and other human-made objects. Topographic and other types of reference maps also show features that do not exist physically in the real world yet provide important context for describing the character of a region. One example is political or administrative boundaries, which may, in part, follow features that do exist in the landscape, such as a coastline or road, as with the recreation area outlined in red in Fig. 13.2. (Note that in some regions, the terms *reference map* and *topographic map* are used interchangeably.)

Reference maps of small areas have a high level of precision, making them appropriate and useful for accurate distance and direction finding, position finding, and navigation. To support these types of uses, some maps that started out as topographic maps may be redesigned to create special-purpose maps. For example, road maps are designed to facilitate navigation over road networks, and hiking maps help readers follow trails on foot. Because of their ability to support a wide range of important map functions, reference maps are used in such diverse areas as engineering, resource management, and city planning.

Reference maps also include very detailed maps of small portions of the earth's surface, such as a construction site or subdivision of city lots. These maps sometimes resemble engineering plans and provide the dimensions and direction of boundary lines and precise areal calculations of land parcels. These types of maps are called plans or plat maps—a plat being a plot of land (Fig. 13.3).

Plans or plats are often the most detailed maps available for an area because they result from data acquired by surveying. A plan whose function is based on delimiting property and landownership boundaries is called a cadastral plan. It provides the basis for legal documents used in land records. Plans and plats are important because other maps of the same area are often derived from them. For example, cities will use these maps to create tax maps, school boundary maps, transit maps, water utility maps, and other products.

13.2.2 Charts

A chart is a special type of reference map whose specific function is to aid navigation and allow readers to determine their position and plot routes or courses. Aeronautical charts display water features; landscape features, including terrain; obstructions to flight, such as towers, smokestacks, and powerlines; and flight-related features, such as airports and airspaces. Nautical charts (Fig. 13.4) show water features and characteristics, including depth, rocks, reefs, and substrate; and aids to marine navigation, such as buoys, beacons, and lighthouses. Charts are often overprinted with navigational tools to assist users in manually plotting a course; for example, a compass rose (shown in the land area in Fig. 13.4) provides the directions of geographic (or true) north, marked by the star, and magnetic north marked by the arrow.

13.2.3 Thematic Maps

Maps that focus on a specific subject or theme are called thematic maps. Unlike reference maps, which show many types of features but emphasize none, thematic maps focus on a single type or characteristic of a feature. While reference maps focus on the variety and relative locations of different features, thematic maps emphasize the geographic distribution of the features or phenomena that relate to the

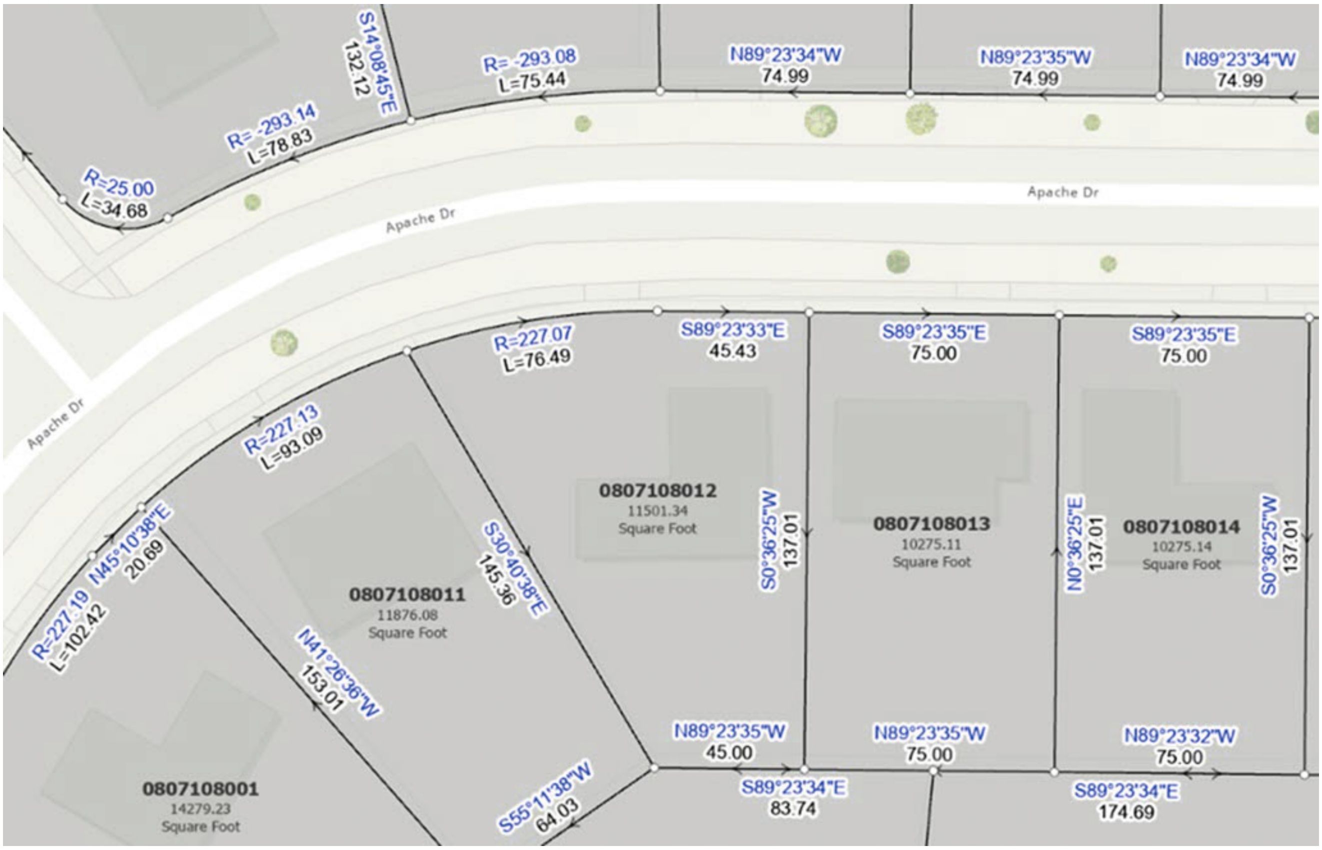


Fig. 13.3 Portion of a tax parcel map in ArcGIS (courtesy of Esri)

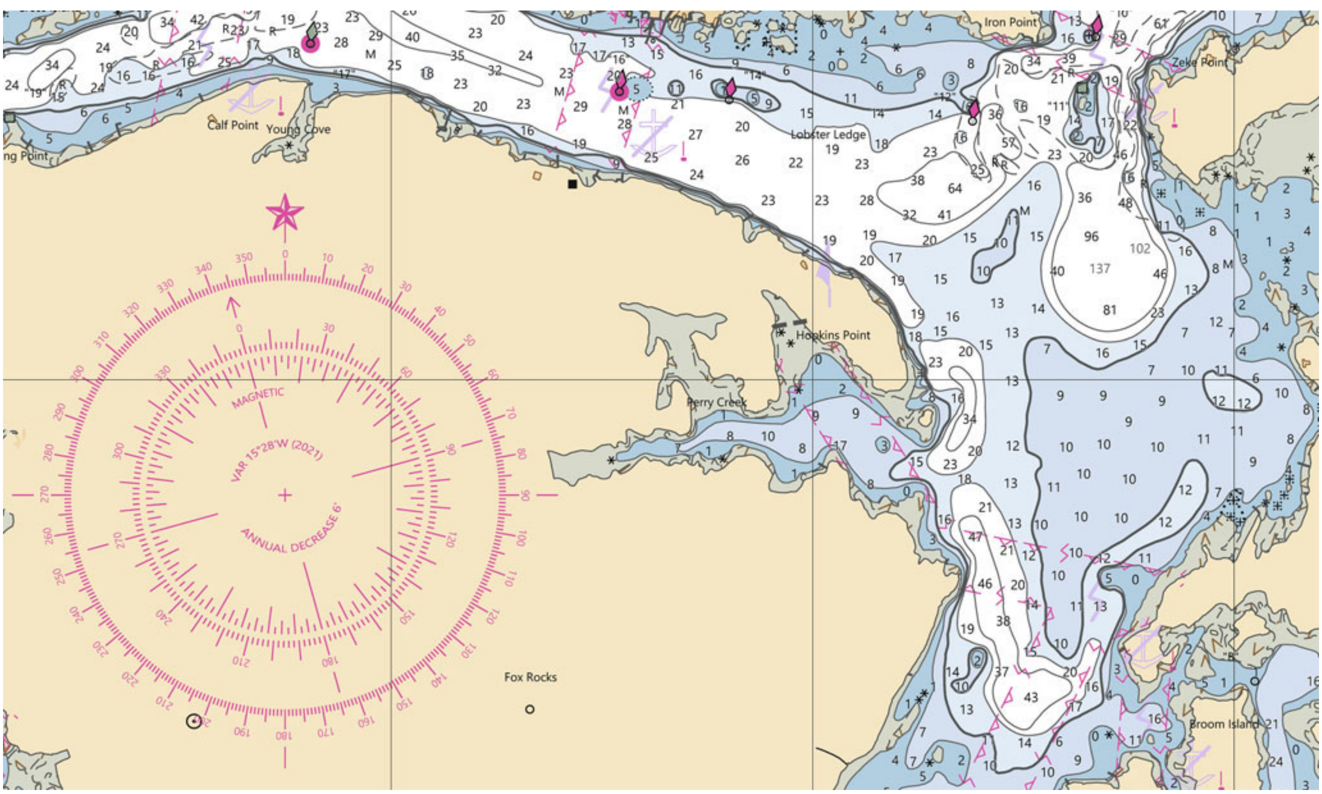


Fig. 13.4 Portion of a nautical chart for the Fox Island Thorofare, Maine, USA (courtesy of National Oceanic and Atmospheric Administration (NOAA), https://www.charts.noaa.gov/OnLineViewer/13308.shtml)

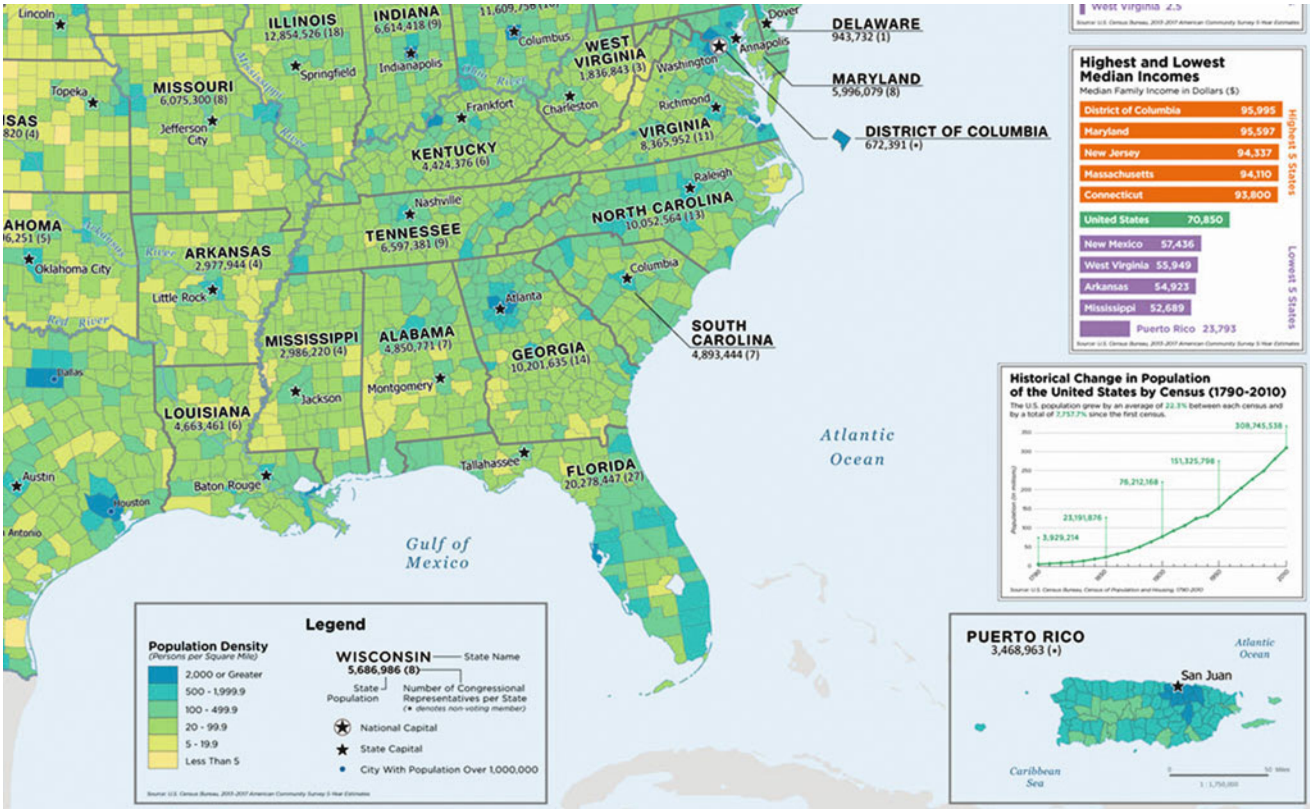


Fig. 13.5 Portion of a thematic map showing population density of the United States (courtesy of US Census Bureau, https://www.census.gov/programs-surveys/sis/2020census/2020-resources/2020-maps/understanding-us-pop-hs.html)

theme of the map, many of which do not physically exist on earth. An example is a map that shows population density (Fig. 13.5). You cannot see this phenomenon in the physical environment, although you may see evidence of it (for example, more houses or apartment buildings).

Thematic maps can depict qualitative information that varies in type but not quantity—such as land use, zoning, and soil classes—or quantitative information based on magnitude—such as population density, annual rainfall, and stream flow.

13.2.4 Qualitative Thematic Maps

Qualitative thematic maps show variation in the kinds of things that are located in the mapped area. This type of map could be used to display the different kinds of trees in a park, the types of crimes that have occurred in a city, or variations in land use. For these maps, it is important that the mapmaker shows the location of features and the variation between feature types, if there is more than one type. Basic reference information, such as political boundaries or city locations, is included to provide geographic context for the reader, but the theme is visually prominent as the most important message of the map. An example is the qualitative map in Fig. 13.6, showing California cities categorized by type with different symbols for the state capital, the county seats, and all other cities. Although the cities themselves provide geographic context, the only other reference information on the map is the county boundaries.

13.2.5 Quantitative Thematic Maps

Quantitative thematic maps show the quantitative attributes (which represent counts, amounts, magnitude, or intensity) of features. Although quantitative thematic maps generally show a single numerical attribute or variable, some are designed to show two (bivariate maps) or more (multivariate maps) variables (Sect. 13.4.2).

Cartographers have developed a variety of standard methods for mapping quantitative data. The appropriate use of these methods depends on the type of information being mapped and the intended use of the map. Choropleth maps, for example, differentiate quantitative attributes of areas through variations in color lightness. These maps can be used when the numerical value within the area is assumed to be homogenous (for example, population density) and when abrupt breaks can be expected to occur between areas (for example, at county boundaries). Figure 13.7 shows a choropleth map represent-

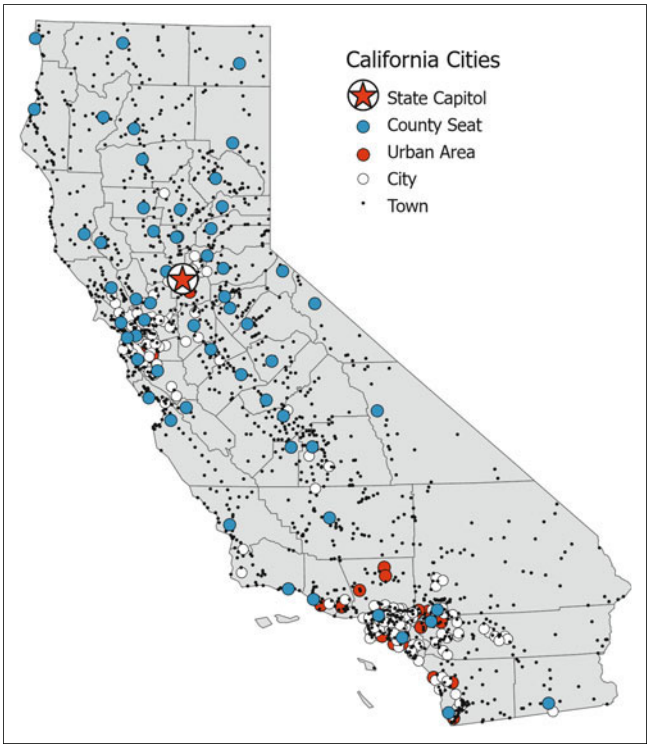


Fig. 13.6 A qualitative map of city types in California (data courtesy of Esri)

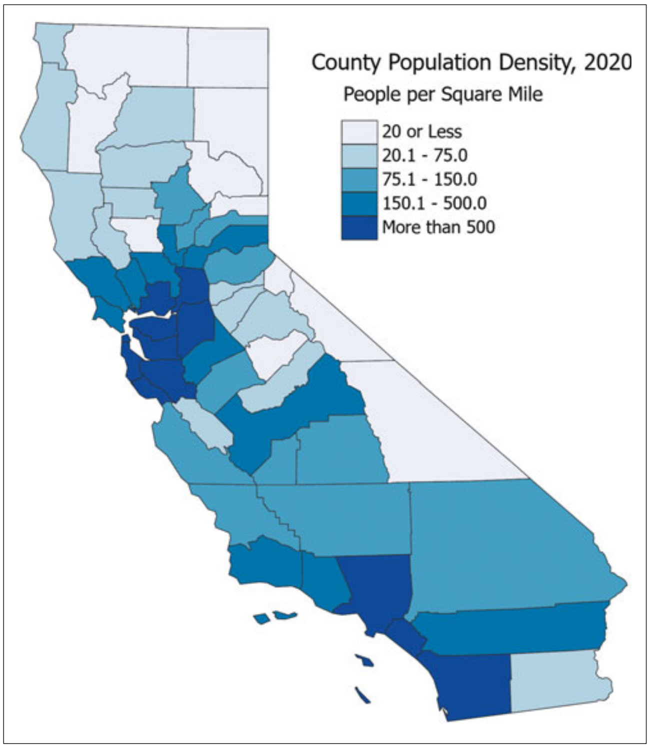


Fig. 13.7 A choropleth map of the 2020 population density for Californian counties (data courtesy of Esri)

ing the population density of California counties. The map is designed to show that some counties have a higher population density (illustrated by darker colors) than other counties (shown with lighter colors). The map also shows areas that display similar characteristics—for example, counties with the same color lightness have the same population density.

An important thing to note about choropleth maps and other maps that assume that the values within the mapped areas are homogenous (for example, Figs. 13.9 and 13.14) is that count (the number of features) and amount (measurable but not countable) data are not appropriate to display with these mapping methods. For example, it would be incorrect to map the raw number of people in a county because all those people could be concentrated within a single area of the county rather than being distributed equally across the county. Notice, for example, that the cities in Fig. 13.8 for the largest county in the state—San Bernardino County, in the southeast—are concentrated in the southwest portion of the county and that there are no cities of significant population size elsewhere in the county. Therefore, when making choropleth maps, it is important to remember to normalize count or amount data by adjusting values so they can be measured on a common scale. Converting population counts to a density (e.g., number of people/area) or a proportion (e.g., number of people in a class/total number of people) are good examples of normalizing data.

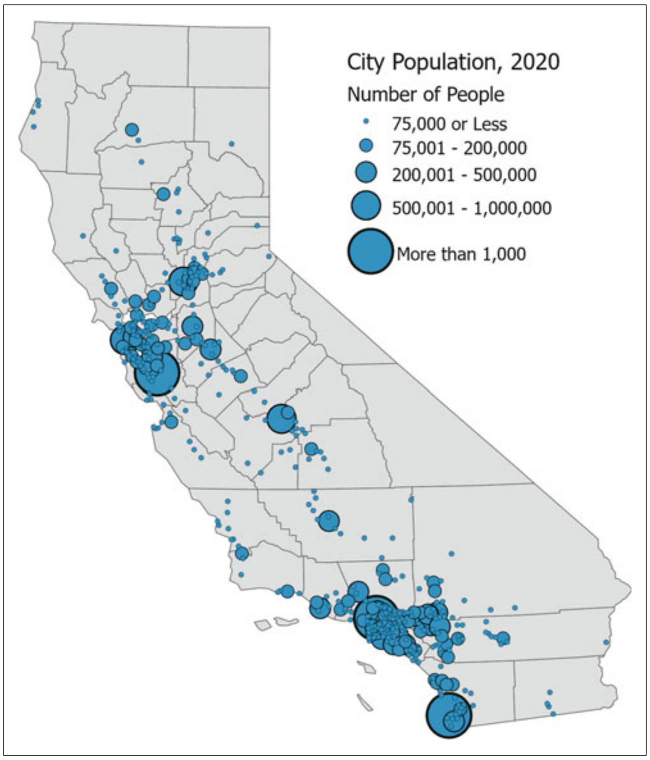


Fig. 13.8 A graduated symbol map of the 2020 number of people in California cities (data courtesy of Esri)

13

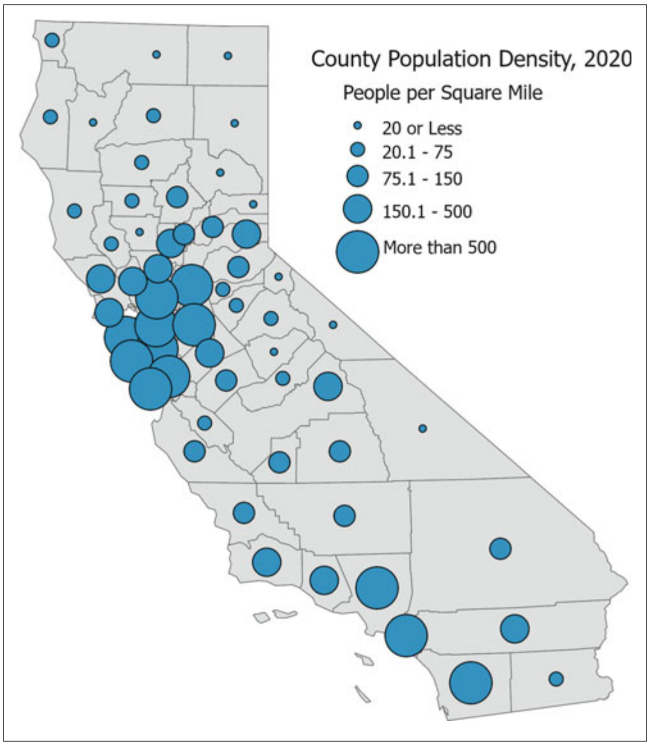


Fig. 13.9 A graduated symbol map of the 2020 population density for California counties (data courtesy of Esri)

A better way to show counts and amounts is provided by graduated or proportional symbol maps, in which the size of the symbol varies relative to the magnitude of the attribute value. Graduated symbol maps divide the range of values into classes, while proportional symbol maps scale the size of symbols in direct relation to the values. These maps can be used to show the values of features at specific locations, such as the number of people in the cities (Fig. 13.8). Strictly speaking, when graduated or proportional symbol maps are used to show values within an area (Fig. 13.9), the data should be normalized for the same reasons that apply to choropleth maps. However, in practice, cartographers will also use these maps to show counts or amounts, as with the proportional symbols used to show number of people in Fig. 13.23.

An isoline is a line connecting points of equal numeric value. For example, the brown lines, called contours, on the USGS topographic map in Fig. 13.2 are isolines of equal elevation, and the gray lines, called bathymetric contours or isobaths, on the nautical chart in Fig. 13.4 are isolines of equal depth. Sometimes, mapmakers color between these lines to show areas with equal values. Isoline maps are conventionally used for portraying data that is sampled at points but that represents a continuous surface, such as elevation, atmospheric pressure, or probability of precipitation (as in Fig. 13.10).

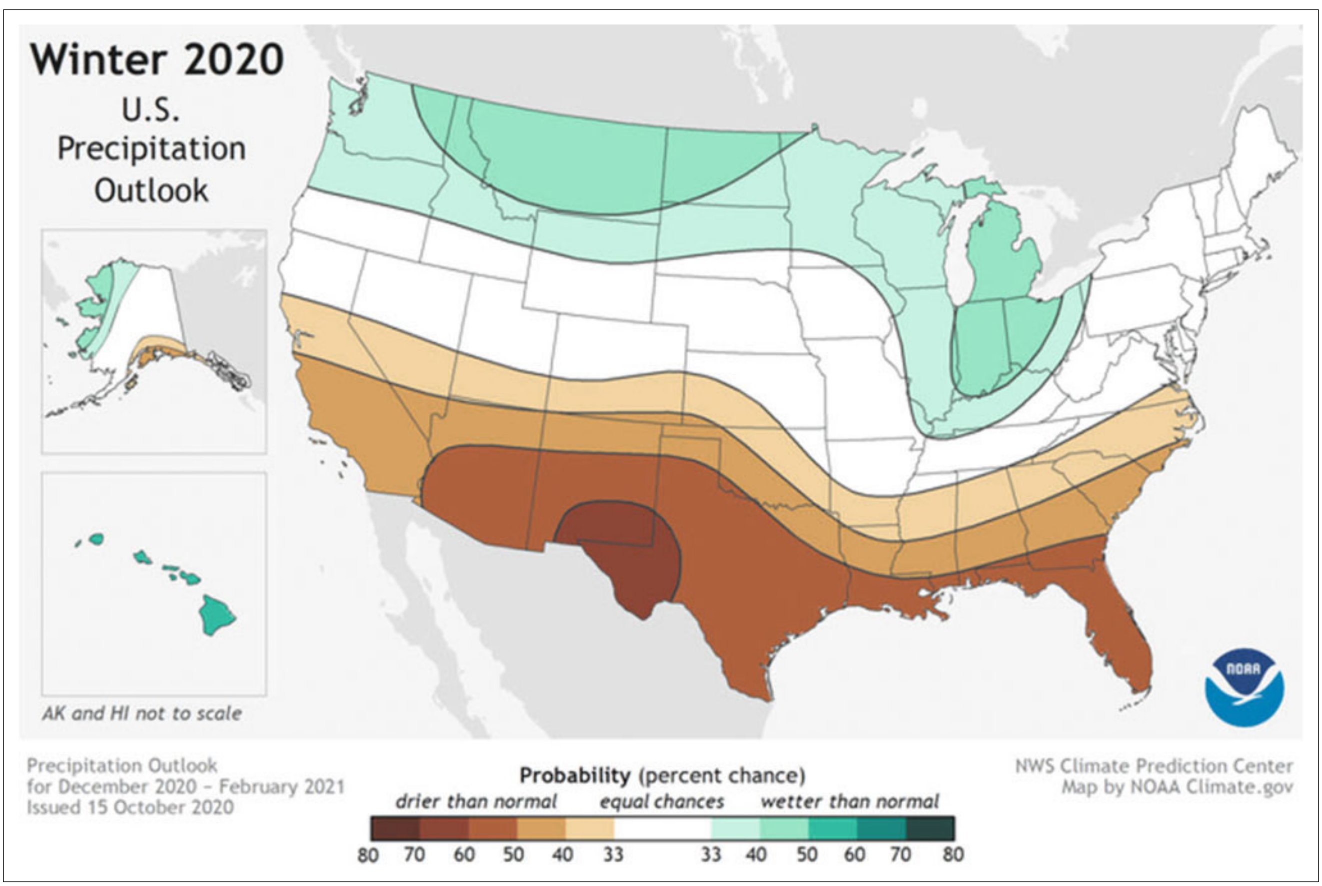


Fig. 13.10 The areas between the isolines are colored to show the precipitation outlook for the United States (courtesy of NOAA, https://www.noaa.gov/media-release/us-winter-outlook-cooler-north-warmer-south-with-ongoing-la-nina)

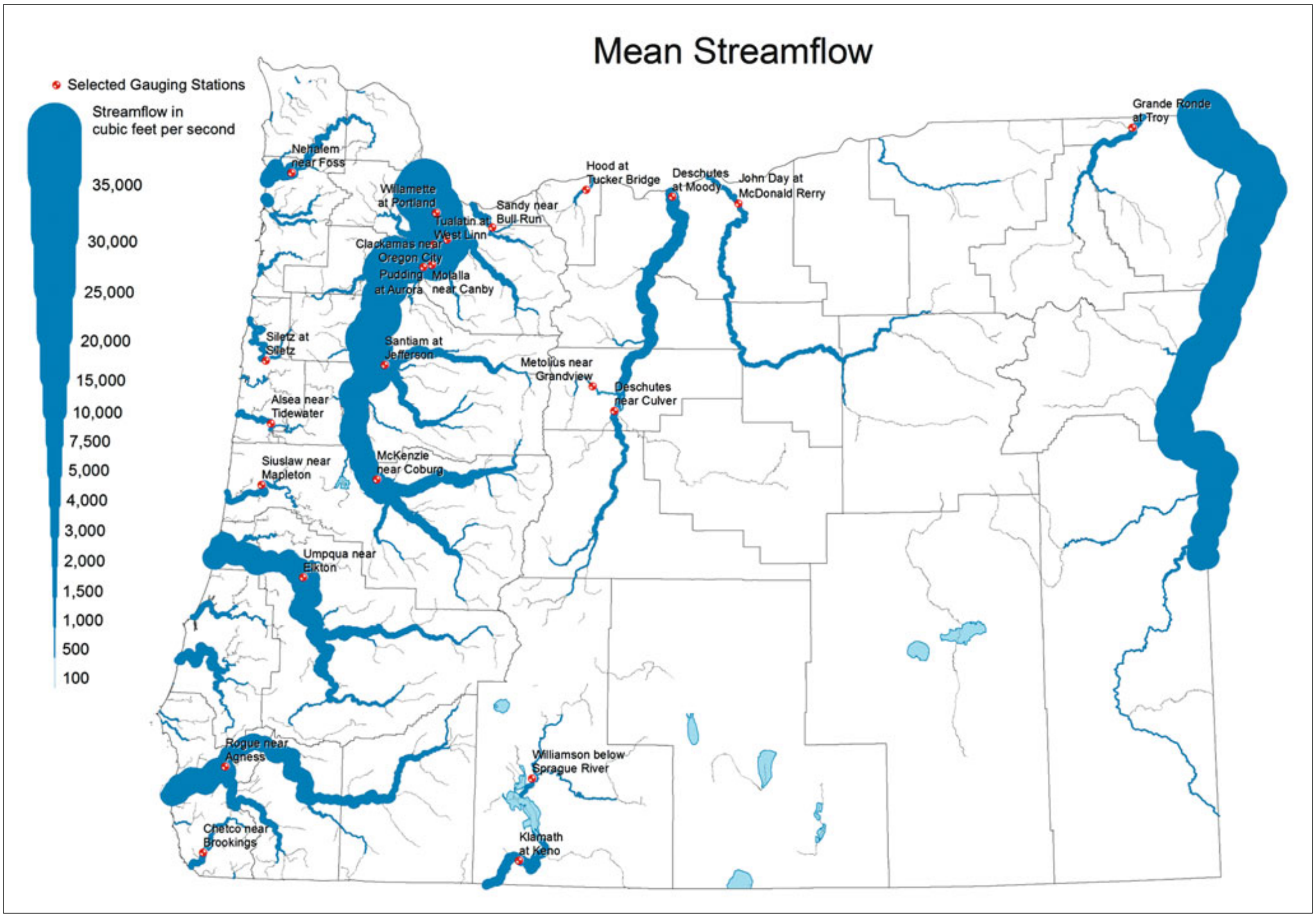
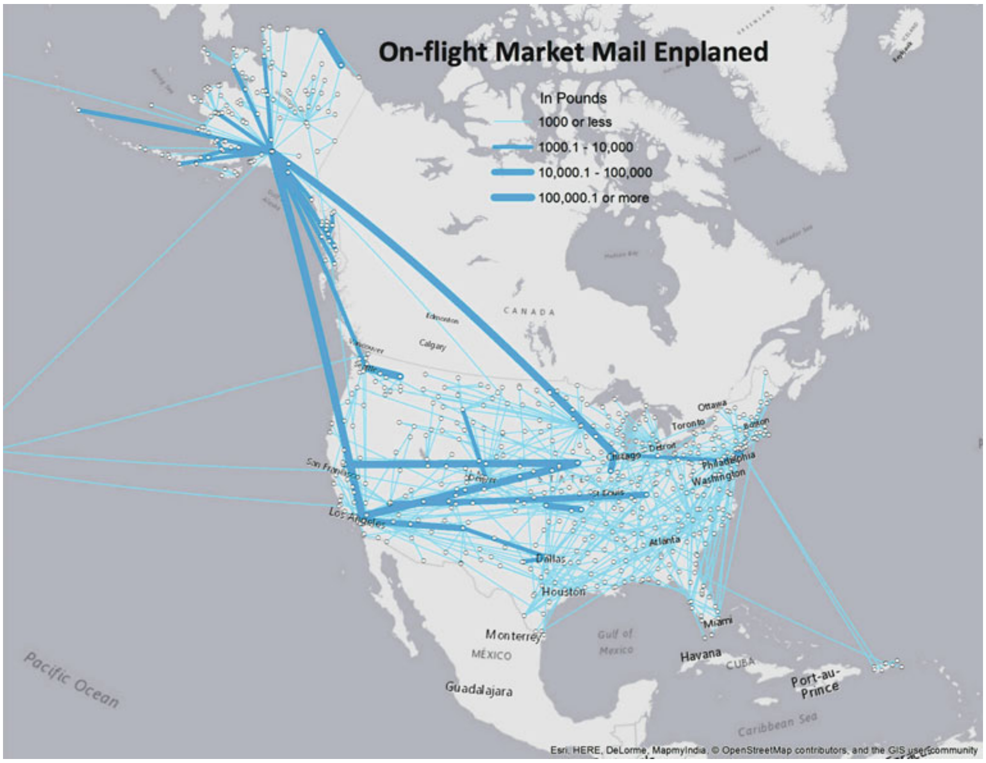


Fig. 13.11 A network flow map showing average annual streamflow for Oregon rivers (data courtesy of USGS and Esri)

Fig. 13.12 A radial flow map showing the amount of mail flown between US airports (data courtesy of Federal Aviation Administration (FAA) and Esri)



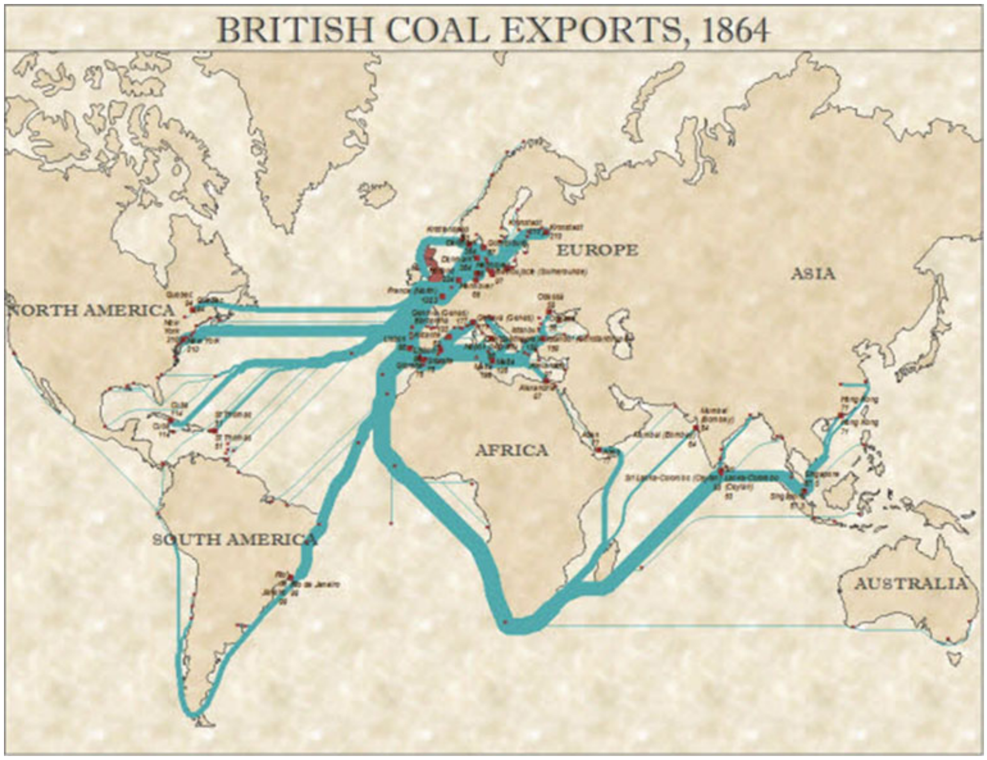


Fig. 13.13 A distributive flow map showing coal exports from Britain in 1864 (data courtesy of Esri)

Another common quantitative map type is a flow map. On these maps, the path or direction of movement is indicated with lines or lines with arrowheads called flow lines. There are three main types of flow maps: network, radial, and distributive. Network flow maps are used to show interconnectivity between places and are usually based on physical networks, such as transportation or hydrographic networks (Fig. 13.11). Radial flow maps have spoke-like patterns because the features and places are mapped in nodal form, often with the shortest route between origins and destinations (Fig. 13.12). Distributive flow maps typically show the distribution of commodities or some other type of phenomena that are delivered from one or a few origins to multiple destinations (Fig. 13.13). The flow lines on these maps may approximate the general supply routes, but they are generalized, unlike the more geographically correct routes on network flow maps (Fig. 13.13).

While there are other methods to map quantitative data, those presented here are the most predominant ones, if we also include dot density maps. Dot density maps—also sometimes called dot maps—show the density of a phenomenon within areas with point symbols (commonly, dots—hence the name of this mapping method). A unique characteristic of dot density maps is that the dots within the areas are shifted to the places where the phenomenon is most likely to occur and excluded from areas where it would not. The density of dots gives the impression of a variation in magnitude (density) of the phenomenon.

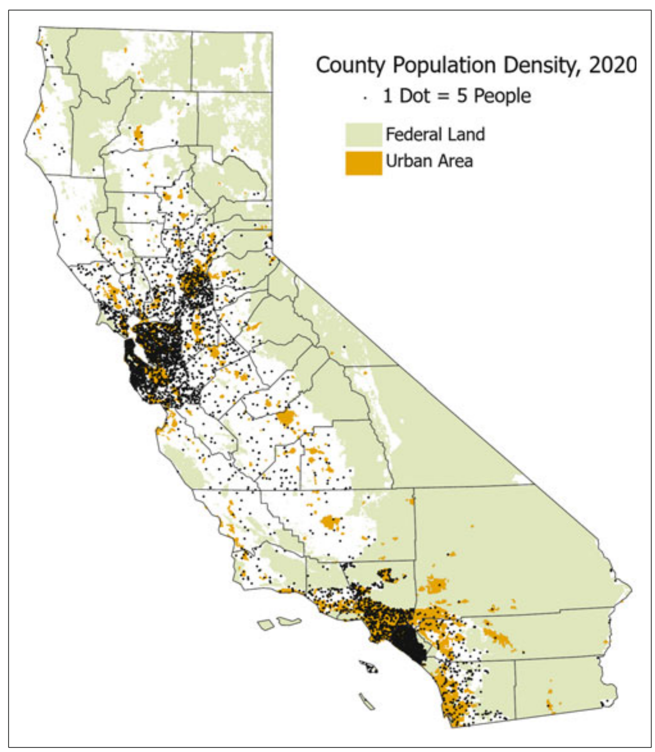


Fig. 13.14 A dot density map of the 2015 population density of California counties, excluding federal lands and including urban areas (data courtesy of Esri)

Dot density mapping is very different from filling a polygon with a pattern of randomly-placed dots, which is actually a form of choropleth mapping because the dots are the fill for the areal symbol. Another important distinction between a dot-filled choropleth map and a dot density map is that the dots on the choropleth map are distributed across the entire area (recall that one of the assumptions for choropleth maps is that the phenomenon being mapped is homogenous across the area), whereas on dot density maps, the dots are excluded from areas where the phenomenon would not exist and concentrated in areas where it would exist. On a dot density map, the dots cannot exist at the same locations as the features because each dot represents more than one feature. The legend on these maps should always indicate the value that each dot represents.

An example of a dot density map is shown in Fig. 13.14. For this map, dots are excluded from the federal lands, shown in green, and included in the urban areas, shown in orange. Notice that the population density in San Bernardino County is represented appropriately, with the dots concentrated in the southwest portion of the county. Dots in the other areas of the county may be located in small areas of public land or pockets of public land within the federal lands.

13.3 Cartographic Compilation

It is sometimes easy to forget that maps are abstract representations of the geographic environment, and people often imbue them with a stronger sense of reality than they should be given. It helps to remember that there are certain properties of maps that result from the cartographic compilation process. That process starts with projecting a scaled down representation of the earth onto a, most often, flat surface. Then cartographers reduce complexity and increase clarity during compilation of the map through selection, generalization, classification, and symbolization of the features on the map. Thus, maps share the following characteristics.

- On all maps, there is a systematic reduction from ground distance to map distance—this is called map scale.
- Except for globes, all maps are made using a map projection, which is a mathematically defined transformation of locations on the spherical (or elliptical) earth to the flat surface used to display the map (for example, a sheet of paper or a computer screen).
- Maps are generalized representations of the environment. Features important to the map are selected and, if appropriate, classified; then they are represented in a generalized form.
- Maps are symbolized representations of the environment. The generalized features are shown graphically, using map symbols and labels.

Because all of these properties are important to consider when a map is being made or used, they are discussed in further detail below.

13.3.1 Map Scale

Map or cartographic scale is the relationship between distances on the map and their corresponding ground distances. Map scale can be expressed as a representative fraction (RF), for example, $1/250\,000$ or $1:250\,000$. In this example, 1 unit of measurement on the map equals 250 000 units of measurement on the ground—for example, 1 inch to 250 000 inch or 1 cm to 250 000 cm. Scale as a ratio is often casually abbreviated to just the denominator—for example, a 250 000 or 250 K map. Map scale can also be expressed as a verbal phrase, such as *1 inch to a mile*, in which 1 inch on the map relates to one mile on the earth. Map scale can also be expressed as a scale bar, which looks like a small ruler on the map or in its margin.

Maps can be classified as either large- or small-scale maps. These terms come from the numeric value of the RF—$1/x$ or $1:x$. For example, the numeric value 1/24 000 is much larger than $1/100\,000\,000$. Thus, the smaller the denominator of the RF, the larger the map scale. A topographic map at a scale of 1 : 24 000 is a large-scale map (Fig. 13.2) in which small areas are shown in great detail, and a world map at a scale of 1 : 100 000 000 is a small-scale map (Fig. 13.13). Note that most of the maps in this chapter are not shown at their original map scale because they have been resized for this publication. This is an important thing to remember about map scale, as the only truly useful indicator for map scale on resized maps is a scale bar (provided it has been resized exactly the same way the map has).

Web maps provide varying levels of detail because additional detail is revealed when users zoom in on an area—these types of web maps are examples of multiscale maps. Each time the user zooms in on the map, the map scale becomes larger. Unfortunately, the map scale is often not indicated on these maps. If this is the case, it is sometimes possible to determine the map scale if the zoom level is known (Table 13.1).

One thing that map scale does for map readers is indicate the appropriate usage of the map. To make accurate distance, direction, and area measurements, use maps at a scale of about 1 : 250 000 or larger. The change in scale across the extent of these maps is negligible, so the map can be trusted as a geometrically correct representation of the small portion of the earth it portrays. If a generalized representation of a large area—such as a state, a country, a continent, or the entire earth—is required, choose a small-scale map. For these maps, remember that scale changes continuously across the map, so if an RF is provided, it only gives the scale at a particular point or along a given line or lines, but not for the entire map. As a result, these maps are not appropriate for calculating measurements of distance or area.

Table 13.1 Map scales associated with the zoom levels for web maps that use the web Mercator projection and tiling scheme (see https://developers.arcgis.com/documentation/mapping-apis-and-services/reference/zoom-levels-and-scale/)

Zoom level	Map scale	Usage suggestion
Level ID:0	591 657 527.59	Global
Level ID:1	295 828 763.80	
Level ID:2	147 914 381.90	Subcontinent
Level ID:3	73 957 190.95	
Level ID:4	36 978 595.47	
Level ID:5	18 489 297.74	Large country
Level ID:6	9 244 648.87	
Level ID:7	4 622 324.43	Small country/US state
Level ID:8	2 311 162.22	
Level ID:9	1 155 581.11	Large metropolitan area
Level ID:10	577 790.55	
Level ID:11	288 895.28	City
Level ID:12	144 447.64	Town
Level ID:13	72 223.82	Village
Level ID:14	36 111.91	
Level ID:15	18 055.95	Small road
Level ID:16	9027.98	Street
Level ID:17	4513.99	
Level ID:18	2256.99	Address
Level ID:19	1128.50	Street intersection
Level ID:20	564.25	
Level ID:21	282.12	
Level ID:22	141.06	
Level ID:23	70.53	

13.3.2 Map Projections

A map projection is a geometric transformation of the earth's surface onto the flat surface used to display the map, most commonly a sheet of paper or a computer screen. This transformation results in distortion of the earth's spherical nature on the flat medium. The map projection process influences the basic geometric properties of the earth's representation, including direction, distance, area, and shape. All maps aim to minimize these distortions or preserve a particular geometric property at the expense of others. The problem is that all the properties cannot be preserved at the same time; therefore, it is important that mapmakers and readers understand how projections maintain or maximize their critical geometric properties and how the other properties are affected.

Many GIS users can and do ignore the map projection issue, particularly if their maps cover a small area not near the poles, because distortion will be minimal on these large-scale maps when the majority of map projections are used. In other cases, GIS users rely on GIS data that is already stored in a projection suitable for the part of the world they are mapping, or they use a projection that has been dictated to them by an authoritative entity or through convention. However, nearly all mapmakers and users will need to know how to make educated projection decisions at some time, particularly for maps of large areas (for example, many countries, the continents, and the world). The challenge is that there are an infinite number of map projections.

So how does the informed GIS user distinguish one projection from another or choose the appropriate one? One way is to organize the wide variety of projections into a limited number of map projection families based on the distortion of the geometric properties discussed above. Therefore, cartographers have developed this set of map projection families:

- Direction (azimuthal)
- Local shape (conformal)
- Area (equal area)
- Distance (equidistant)
- Shortest route (gnomonic).

Some map projections retain more than one of the geometric properties. For example, an equidistant azimuthal map projection preserves both direction and distance. However, no map projection can simultaneously preserve shape and size. To learn more about map projections and their distortion properties, see standard cartography textbooks, such as [2], Snyder's classic book [3], or web resources, such as those listed at http://en.wikipedia.org/wiki/List_of_map_projections.

Despite the infinite number of projections to choose from, in practice there are a relatively small number of projections that are regularly used because they are appropriate for common combinations of extent, map scale, and intended map use. A resource that summarizes the commonly used map projections is the USGS map projection poster (https://www.usgs.gov/media/files/map-projections-poster).

Although there is a wealth of information on map projections, all too often map projections are used incorrectly. A good example is the Mercator projection (Fig. 13.15), which is unique in that any line drawn on a map using this projection is a line of constant direction or bearing. This property makes the projection good for navigational maps and charts, since navigators can draw a line from their current position to their destination and simply set the compass bearing in that direction. If they keep following the bearing (not accounting for air or water currents and the like), they will eventually arrive at their destination. In practice, the process is more complex than this, and navigators will first find the line of the shortest route (which can be done using a gnomonic projection), then convert that to a series of shorter straight lines on a Mercator projection map, which will be used to set and reset their course.

Although constant direction is an important and desirable property for maps that support navigation, it is countered by great distortion in shape and size (Fig. 13.15), which makes the Mercator projection inappropriate for almost any other use. Nonetheless, for decades, wall maps in children's classrooms were made with this projection. These maps would be a better teaching aid if they were made with a more ap-

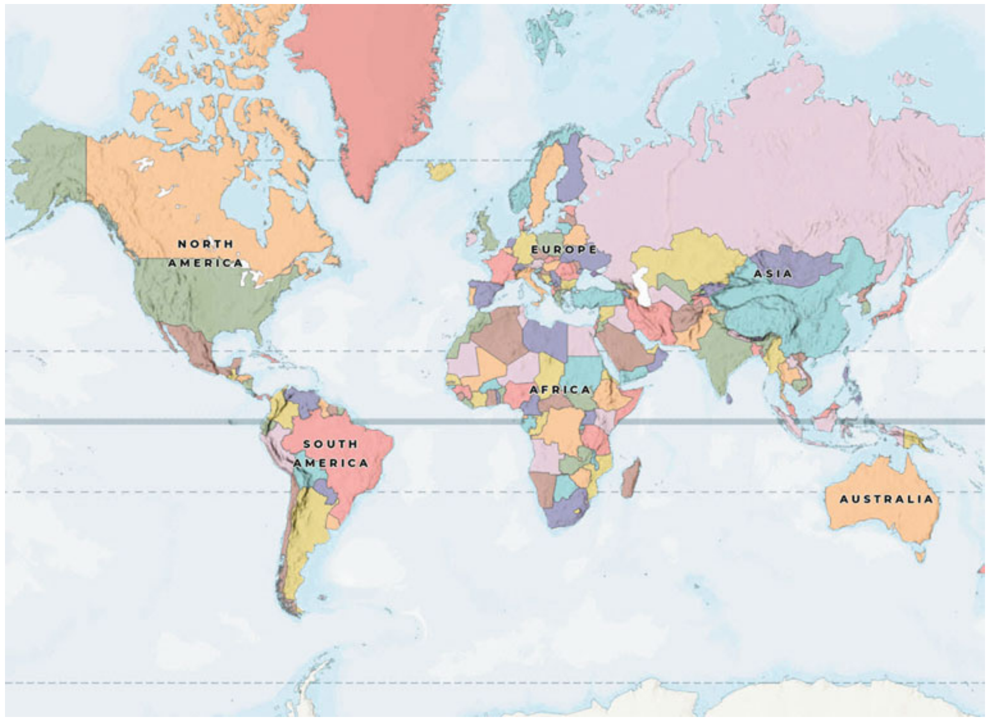


Fig. 13.15 World countries shown on a map with the Mercator projection (data courtesy of Esri)

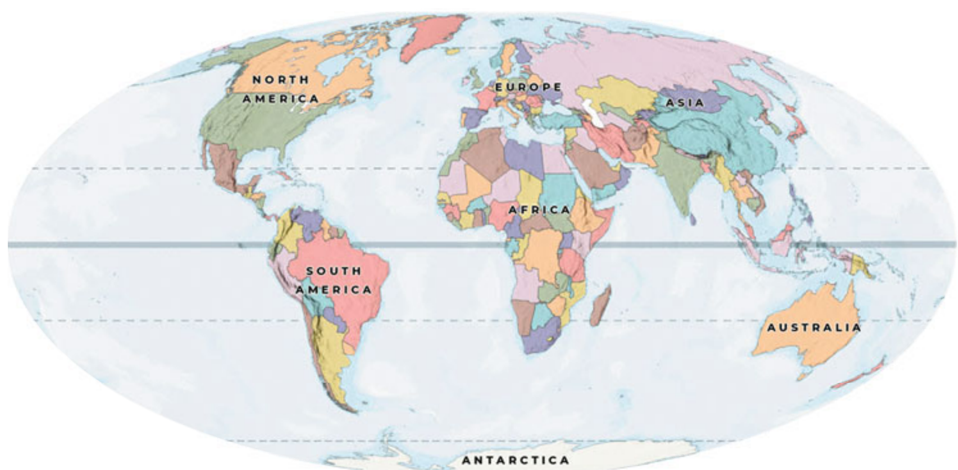


Fig. 13.16 World countries shown on a map with the Mollweide projection (data courtesy of Esri)

propriate projection, for example, the Mollweide projection (Fig. 13.16), which is an equal area map projection used for the world.

A modified web Mercator projection has become the de facto standard for web maps. (See [4] for an in-depth examination of the web Mercator projection.) While this is fine for viewing small areas (zoom levels 11 through 23, as shown in Table 13.1) that are not near the poles, it is inappropriate for most smaller scale maps of continents and the world. Nonetheless, the web Mercator projection has found widespread use, and examples of world maps made with it abound. As a result, generations of map users, including users of modern web maps, may have a distorted mental view of the relative size and shape of the world's land masses.

13.3.3 Selection

The first step in abstracting information about the world into something that can be represented on a map requires selection—deciding what type of and how much information to portray. This selection of information reduces the complexity of the world to make the map more intelligible. Selection not only increases clarity but also helps the map reader understand what is important on the map.

A map that will be used for a range of purposes, such as a topographic map, must include a lot of information of various types, although only some of it may be relevant for particular map uses. For these maps, selection is a more arduous task. For thematic maps, only the locational information that is required to provide geographic context and is relevant to the theme should be included, because superfluous information can result in distraction from or misinterpretation of the map's message. For thematic maps, the challenge is to avoid adding too much superfluous content.

13.3.4 Generalization

Once the features important to the map have been selected, they are then represented in a generalized form. Cartographic generalization reduces the amount of information on a map through a change in the geometric representation of features. Common generalization operations for vector data are shown in Fig. 13.17. A generalization operation may be applied to a single feature or a single type of feature, but the results of that operation may affect other features or categories of features. For example, smoothing a line symbol that depicts a river may result in the generalized line overlapping streamside buildings or roads or being misaligned with bridges and contour lines. As a consequence, an additional generalization operation—for example, displacement—may be used. Most generalization tasks, therefore, require that the results be considered contextually (relative to other features) to ensure graphic clarity and allow inherent geographic relationships and patterns to be recognized—this is known as contextual generalization. It is easy to understand why multiple generalization operations are often used. The most important consideration for a cartographer is that, whether applied to a single feature or a collection of features, the outcome of any generalization operation or set of operations should always preserve the characteristic appearance and nature of the local geography.

Raster data can also be generalized. For example, the satellite image in Fig. 13.18 can be classified to show land cover (Fig. 13.19). A filter can be used to replace the values of isolated cell with those of the majority of their contiguous neighboring cells. A boundary smoothing operation can be used to simplify the boundaries between classes. Pockets of cells that do not meet a size threshold can be grouped into larger regions. These are but a few examples of raster generalization operations, the goal of which is to produce a more generalized map with larger groups of cells that have similar values (Fig. 13.20).

13.3.5 Classification

Selected and generalized data can be further simplified through classification. Classification is the grouping, ordering, or scaling of features into categories (for qualitative data) or classes (for quantitative data). This is usually done based on attributes of the features, although sometimes it can be done by some other criteria, such as using the location of features to create clusters.

For quantitative data, there are several commonly used classification methods (or classing methods). Classification methods are procedures used to assign class intervals or ranges to numerical distributions. The selection of class intervals and the number of classes both have a significant impact on the appearance of the quantitative information on a map. Although class intervals can be set manually, most GIS systems provide a set of common classification methods that users can choose from, three of which are described below.

With quantile intervals, the same number of features is assigned to each class. If the number of classes is four, the classes are quartiles, five classes are quintiles, and so on. Data can also be classified by equal range intervals, commonly called equal intervals, in which the range of data values is divided by the desired number of classes. Another classification method that is commonly used in GIS is to specify classes based on breaks between groups in the distribution of data values. With natural-break intervals, the goal is to minimize the variation within classes and maximize the variation between classes. This can be done using a method described by Jenks [6]—thus, this is often called the Jenks natural-breaks classification method or Jenks optimization method. For a complete description and illustrations of these methods and others, see [2].

13.4 Symbols

The final step in the cartographic abstraction process is symbolization, in which features and their attributes are represented by graphic elements, called symbols. Symbols do not always take on the appearance of the geographic feature they represent, for example, a simple circle does not look like the city that it represents on a map (see Fig. 13.6). Nonetheless, the map reader should be able to interpret the symbol's meaning accurately and relate it to its real-world counterpart appropriately.

13.4.1 Properties of Symbols

Cartographers commonly consider three properties of data that guide their design and use of the symbols used to represent the data dimensionality, level of measurement, and visual variables [8]. Dimensionality is an expression of the spatial extent—especially width, height, or length—of a feature. For cartographic purposes, dimensionality relates to the following types of symbols:

- Point symbols (zero dimensions)
- Line symbols (one dimension)
- Area symbols (two dimensions)
- Surface representations (2.5 dimensions; Sect. 13.7)
- Volume representations (three dimensions).

The dimensionality of the feature in the real world may be the same as the dimensionality of the feature's symbol on the

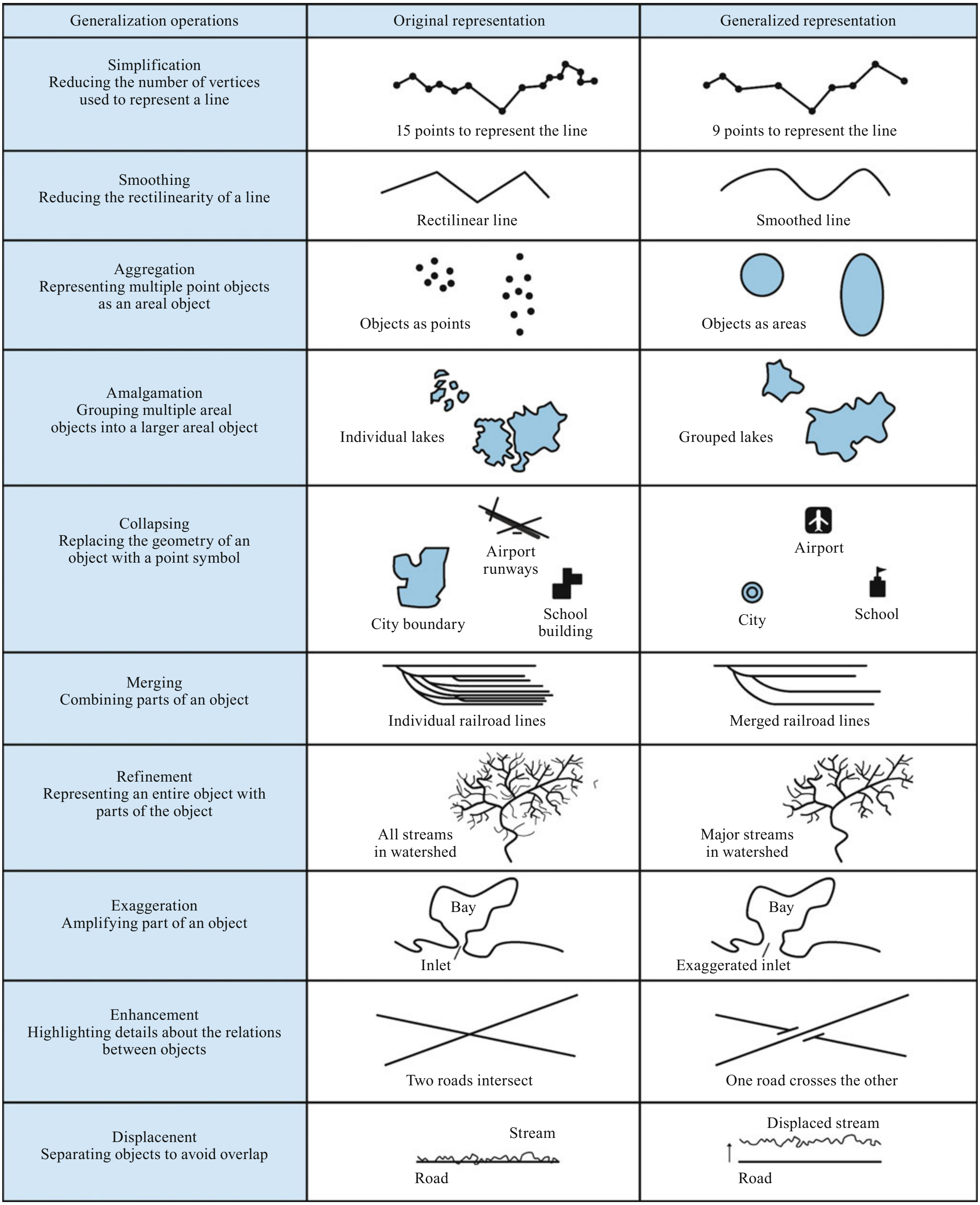


Fig. 13.17 Common generalization operations for vector data (adapted from [5])

Fig. 13.18 A small portion of 30-meter Landsat 8 imagery (bands 6,5,2) focused on the area around Plaster Rock, New Brunswick, Canada (data courtesy of USGS and Esri)

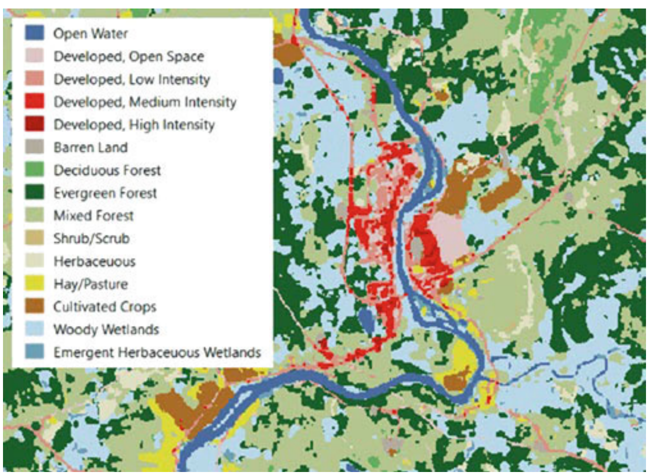


Fig. 13.19 A portion of the land cover map after classifying the much larger satellite image (data courtesy of USGS and Esri)

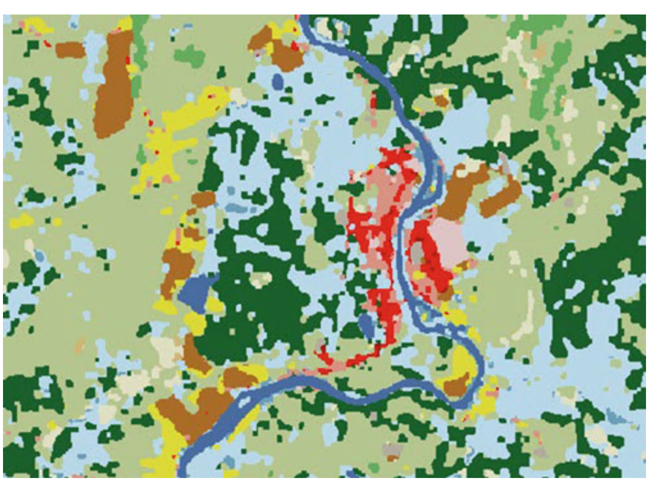


Fig. 13.20 The land cover map after raster generalization operations have been applied (data courtesy of USGS and Esri)

map. For example, an object that exists as a linear feature in the geographic environment, such as a powerline, may be mapped using a line symbol, such as a black line; a vegetated area may be mapped using a green polygon; and a surveyed location may be symbolized with a brown cross. The correspondence between the dimensionality of a feature and its symbolized representation is not always so straightforward. This is due to the other factors that relate to the properties of maps, including map scale, map projections, and generalization, but foremost among them is map scale. As map scale changes, the spatial dimension of a mapped feature may also change. For example, on a large-scale map, a city may be represented as an area with two dimensions. However, the same city might be better represented as a point feature with zero dimensions on a small-scale map (as with the state capitols in Fig. 13.5), because the space available to show the city is restricted.

Another property of symbols relates to the level of measurement of the data they are being used to represent. For cartographic purposes, it is often sufficient to classify the data as either qualitative or quantitative. Thematic maps showing these types of data were discussed in Sects. 13.2.4 and 13.2.5.

The third property of symbols relates to how their appearance can be altered or varied. In cartography, the graphic variations of map symbols are called visual variables, a concept introduced by Bertin in 1967 [7]. Visual variables provide mapmakers with a range of possibilities to consider when designing symbols and assigning them to features on maps. Visual variables can be categorized as those that differentiate and those that order. Differentiating visual variables are used to show qualitative differences among features, while ordering visual variables show quantitative differences. Cartographers refer to these as the qualitative and quantitative visual variables, respectively.

The qualitative visual variables include shape, orientation, and color hue (Fig. 13.21).

- The shape used for point symbols can range from simple geometric forms (circles, squares, or triangles) to miniature representations that mimic the feature (called mimetic symbols), such as a pictorial sketch of a landmark building, as shown in Fig. 13.21. Shape, in reference to line symbols, relates to the marks that make up the line. For example, different combinations of solid or dashed lines and point symbols are used to portray different kinds of linear features. As with line symbols, shape for area symbols relates to the marks that make up the symbol's fill. The marks in line and area symbols also range from simple geometric shapes to detailed mimetic symbols.
- Color hue refers to the dominant wavelength of visible light that forms our color association (see Sect. 13.5.1). What we commonly recognize as the colors red, blue, and yellow are more accurately referred to as hues. Strictly

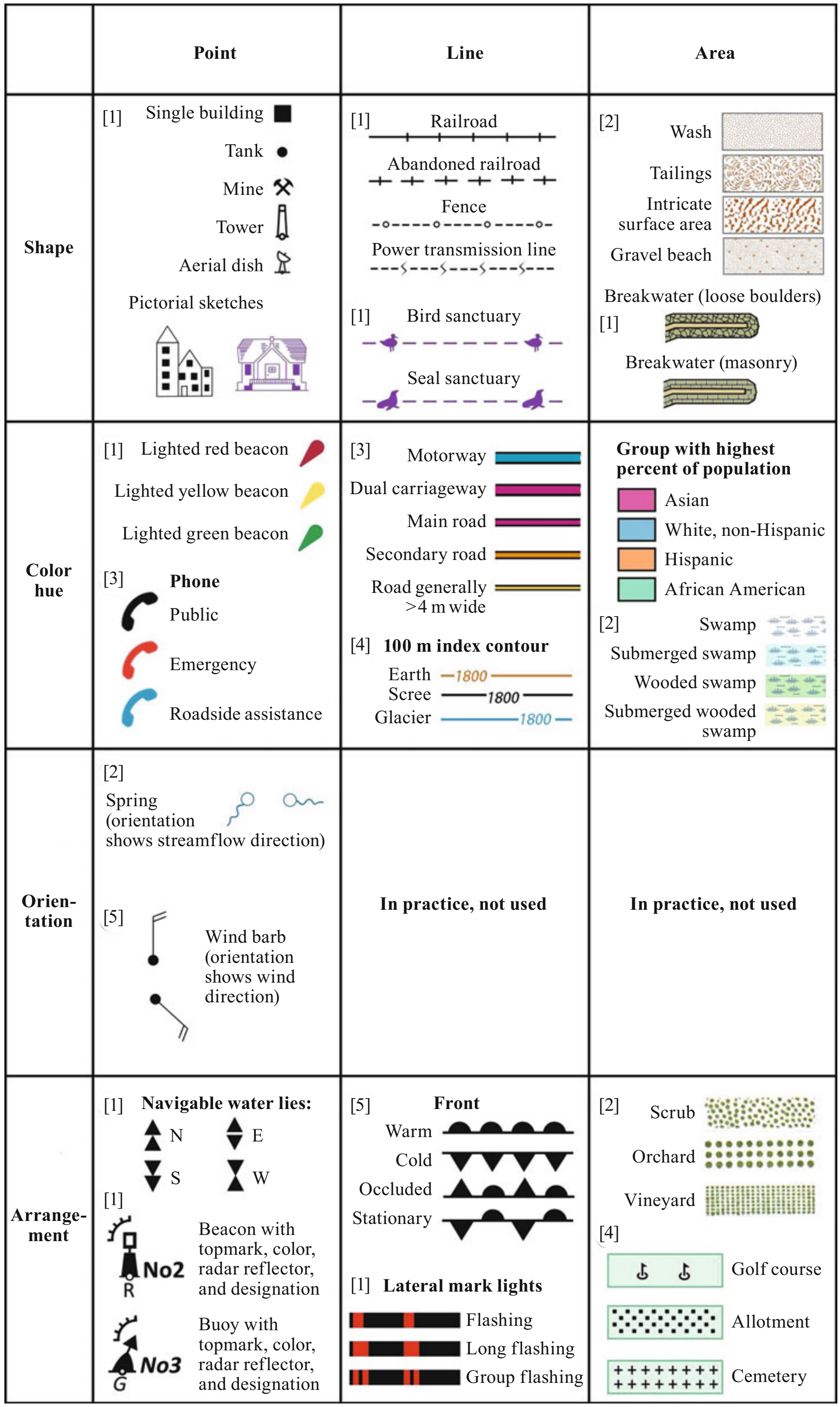


Fig. 13.21 Qualitative visual variables (adapted from [2] NOAA nautical charts, [3] USGS 1 : 24 000-scale topographic maps, [4] Ordnance Survey 1 : 25 000-scale Explorer maps, [5] swisstopo national maps, and [6] NOAA weather maps)

speaking, variation in hue should not impart a message of magnitude, as red is not more than blue, which is not more than yellow. However, some colors have an inherent lightness (for example, yellow) or darkness (for example, purple), so there may be some internal ordering by the human visual system. Additionally, some hues have an inherited meaning through common usage. For example, on weather maps, red is often associated with warmer temperatures and blue with cooler temperatures. These considerations should be kept in mind when choosing hues for qualitative map symbols or reading color maps.

- Orientation relates to the angle of the symbol (or a portion of the symbol) for point features, as with the spring symbols in Fig. 13.21. In common practice, orientation is not used for line or area symbols since the orientation of the corresponding features is already determined by their geography [7]. Although orientation can be measured in angles from 0–360°, this numeric value does not impart a magnitude message because north is not more than east, which is not more than south.
- Arrangement relates to the positioning of marks within a symbol. An example is a random arrangement of green circles to denote an area with scrub or a regular array of similar symbols to show orchards and vineyards, as illustrated in Fig. 13.21.

The quantitative visual variables include size, color lightness, color saturation, and pattern texture (Fig. 13.22).

- The size of a symbol can be varied, which both the point and line examples in Fig. 13.22 illustrate. The visual impression of point and line symbols varying in size is one of a measurable or ordered difference in the magnitude of the attribute being symbolized. In common practice, size is not used for area symbols, since the area of a feature is already dictated by its geography [7].
- Color lightness, also called color value, refers to the darkness or lightness of a color (see Sect. 13.5.1). The human visual system will naturally interpret a darker color as *more*, so darker colors are used to represent features with greater quantities or higher quantitative data values than features shown with lighter symbols. In Fig. 13.22, the color lightness of the point symbols for each of the race and ethnicity categories (distinguished by color hue) represents the percentage of population.
- Color saturation, also referred to as color chroma, is the purity of a hue (see Sect. 13.5.1). Saturation is varied by mixing a hue of high purity with a gray of the same lightness—the addition of gray results in lower saturation. The terms *brilliant* or *intense* can be used to describe fully saturated colors, and *muddy* or *dull* describe colors with low saturation. In Fig. 13.22, color saturation is used to show percent slope for each of the aspect categories, which are differentiated by hue. In practice, color saturation and color lightness are often used together to create sets of colors ranging from light to dark that can be used to symbolize quantitative data (Sect. 13.5.6).
- As with the shape and arrangement of qualitative visual variables for area symbols, pattern texture relates to the marks that make up the symbol. Denser symbols with less space between the marks will appear to be darker. Thus, pattern texture acts in the same way as the color value quantitative visual variable so that denser patterns denote higher values, as illustrated in Fig. 13.22.

13.4.2 Bivariate and Multivariate Symbols

Mapmakers sometimes combine two or more visual variables to emphasize a single attribute or to show multiple attributes. To emphasize a single quantitative attribute, mapmakers can use a primary visual variable, such as size, augmented by a secondary visual variable, such as color value (lightness), in which darkness serves to exaggerate the message that is conveyed by a larger size. Cartographers might also combine two visual variables to display two different attributes. Figure 13.23 provides an example in which color hue is used to differentiate the year and symbol size is used to express the number of people. Mapping two variables is achieved with a bivariate symbol, as with the circle symbols in Fig. 13.23; mapping more variables requires a multivariate symbol, such as the beacon and buoy symbols in Fig. 13.21. An example of a bivariate choropleth map is shown in Fig. 13.24. On this map, color hue differentiates the two categories of attributes (cancer mortality for males and females), and color value portrays the quantitative value of the attributes in the categories. Purple hues denote areas where the values for both males and females are similar.

Of course, the more complex the symbol, the harder it is for the map user to read the map; the longer it takes to understand the map; the greater the chance the map will be misinterpreted; and the harder it is to design the symbol (because the guidelines for each visual variable must be correctly applied). Therefore, the need for bivariate and, especially, multivariate mapping should be well justified.

13.4.3 Design Guidelines for Symbols

The range of possibilities when designing cartographic symbols is virtually limitless. While this can seem daunting to the budding mapmaker, seasoned cartographers base their decisions on the conceptual foundation for symbol design described in Sect. 13.4.1—that is, the triad of dimensionality, levels of measurement, and visual variables. The examples in Figs. 13.21 and 13.22 illustrate how these three concepts come together. Notice that the levels of measurement are reflected in the two sets of examples—qualitative visual variables in Fig. 13.21 and quantitative visual variables in Fig. 13.22. Also notice that the dimensions are represented in the columns of the two figures, and the visual variables are represented in the rows. (Symbols for 3-D features are not discussed in this chapter.) For the most part, when the three concepts are considered carefully and in concert, cartographic symbol design is controlled, confined, and

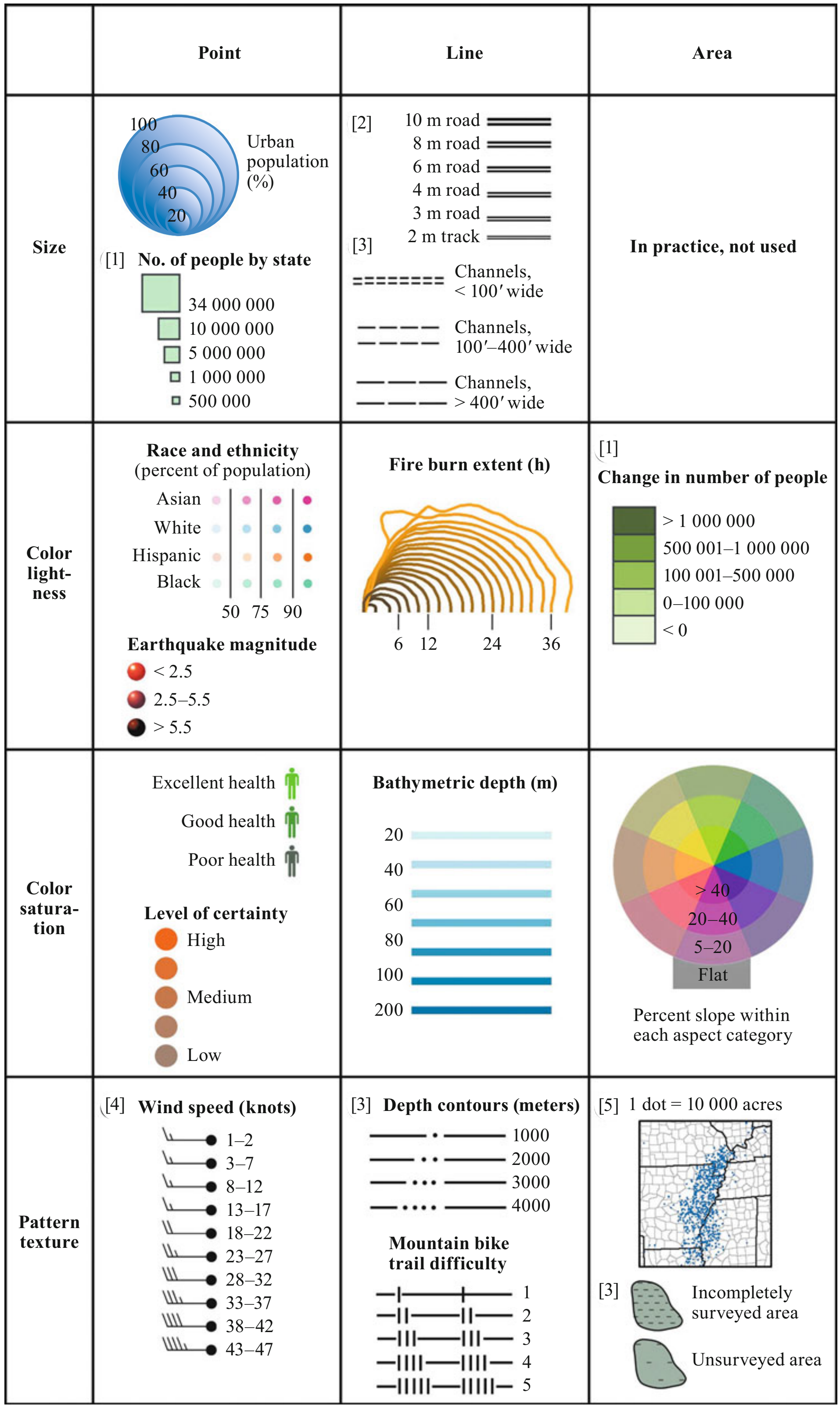


Fig. 13.22 Quantitative visual variables (adapted from US Census Bureau atlas maps [2], swisstopo national maps [3], NOAA nautical charts [4], NOAA weather maps [5], and US Department of Agriculture (USDA) agricultural atlas maps [6])

often conventional, although there is still room for creativity. Many symbol design guidelines have been provided earlier in Sect. 13.4; others are provided in the discussions of color (Sect. 13.5) and type (Sect. 13.6).

Sometimes map symbols can be interpreted without the use of a legend (Sect. 13.9.1 gives more details on legends). This can be achieved by using symbols that are familiar because of cartographic conventions (such as a blue line for

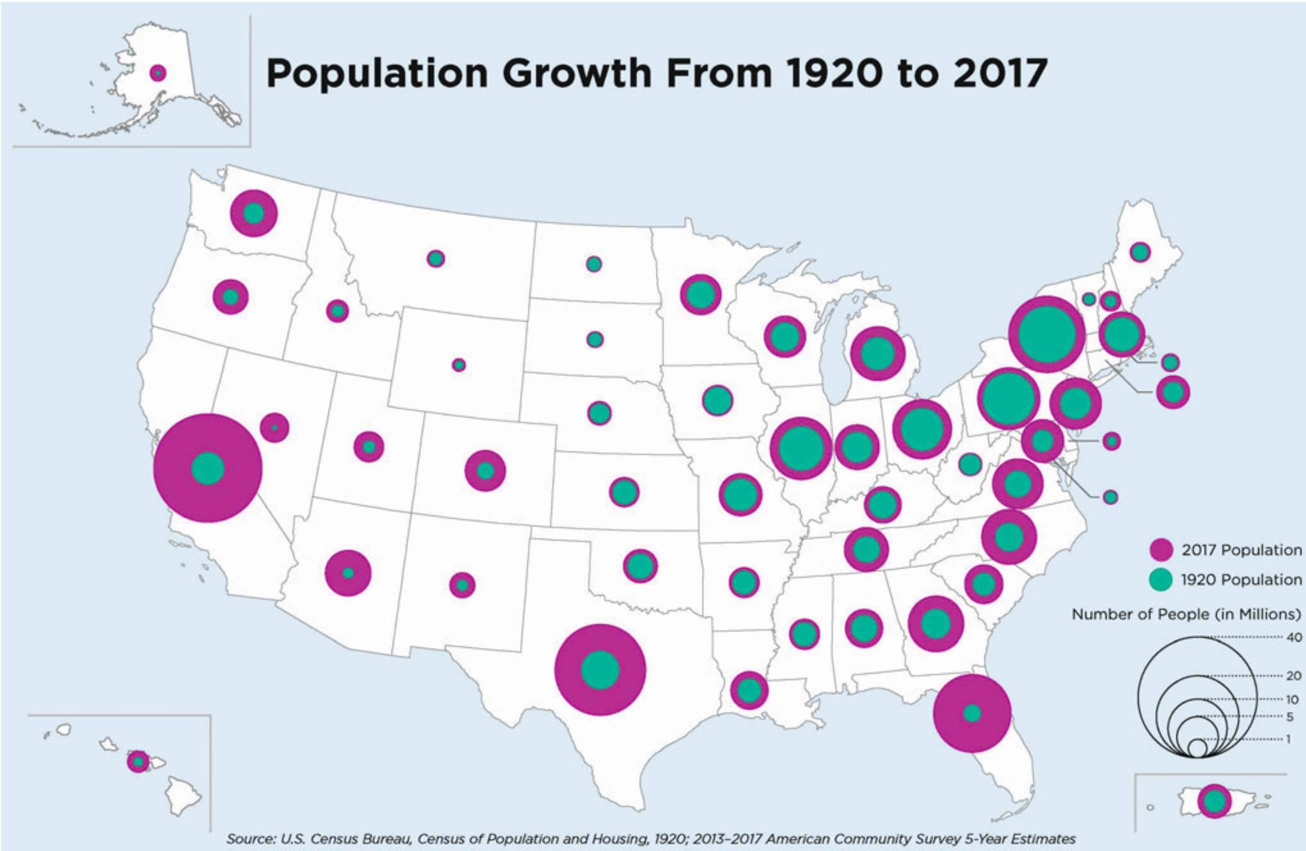
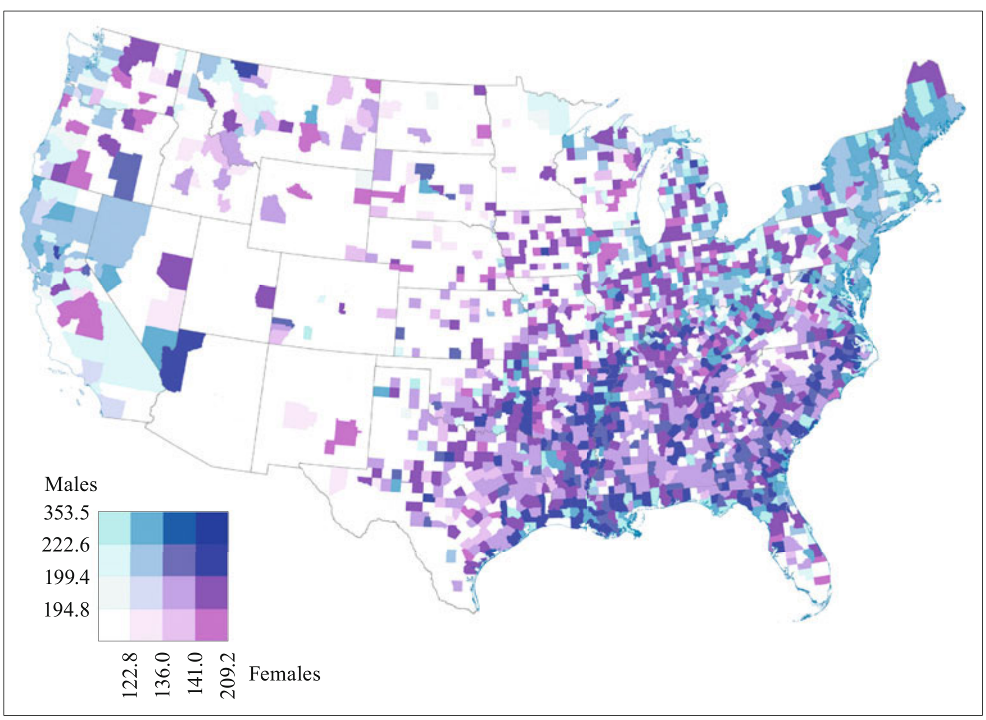


Fig. 13.23 A bivariate symbol map for US counties showing population growth from 1920–2017 (courtesy of US Census Bureau, https://www.census.gov/programs-surveys/sis/2020census/2020-resources/2020-maps/understanding-us-pop-hs.html)

Fig. 13.24 A bivariate choropleth map for US counties showing 1970–1994 cancer mortality rates for both males and females (data courtesy of National Cancer Institute (NCI) and Centers for Disease Control and Prevention (CDC))



a river), using unambiguous symbols (for example, using a tent symbol for a campsite), or labeling features directly on the map (as with a trail or railroad). If any opportunity for misinterpretation exists, a legend should accompany the map, and the symbols in the legend should appear exactly as they do on the map. A telltale sign of a novice mapmaker is a legend with symbols or colors that are different from those on the map (for example, when a point symbol in the legend is larger or smaller than the corresponding symbol on the map or when transparency that is used on the map is not reflected in the colors in the legend).

13.5 Color

Color not only makes maps more interesting to look at, but it can also be a great aid to mapmakers, since it allows more types of features to be distinguished on the map and better order to be imposed within the map (Sect. 13.8). One of the major challenges for mapmakers is to use color in an effective manner so that it aids in communication rather than complicating and confusing the message. Today's computer systems with 24-bit color depth can display 16.7 million colors (32-bit color depth with an alpha channel supports almost 4.3 trillion color combinations), making the challenge of color selection even more difficult. Mapmakers, therefore, benefit from basic knowledge about the properties of color, familiarity with common color models, and guidelines on how to use color on maps, some of which were described in Sect. 13.4.1.

13.5.1 Properties of Color

Although one person's conception of blue may be slightly different than another person's, most people with normal color vision do not differ substantially in their color perception. Therefore, maps can usually be safely designed based on the assumption that the map user has normal color perception. There are exceptions when designing maps for specific conditions of use (for example, to be read in low light) or by specific audiences (such as readers with color vision deficiency).

Several systems based on observers with normal color vision have been developed to describe color properties and define colors. These systems have to be more detailed than simply assigning a specific color to the label *blue*, for example, since the range of possible blues renders the description meaningless. Furthermore, in English, there are only 11 basic color terms: black, white, gray, red, green, yellow, blue, brown, purple, orange, and pink (see Fig. 13.25, top). Consequently, color descriptions must be more precise. This is possible because we can describe color based on the physical stimulation of color and the perception associated with viewing colors. Being able to define color with precision is important in order to control color throughout the mapping process, from data input to graphic display to map output. For print maps, for example, the process involves selecting a color (often from a color chart) and ensuring that the color's settings are used throughout the entire design process. Because a particular color has been specified, the mapmaker can be confident it will look correct when printed on paper or displayed on a computer screen.

Color specifications are often based on the three basic elements of color (Fig. 13.25):

- Hue: commonly referred to as color, such as red, green, or blue.
- Lightness: the lightness or darkness of a hue; also called value. Lighter colors are sometimes referred to as tints—the mixture of a color with white; darker colors are called shades—the mixture of a color with black.
- Saturation: the extent to which a color departs from being neutral, such as gray, or in simpler terms, its *colorfulness*.

These three elements of color must be considered relative to the medium used to display the map and the conditions of map use. For example, color lightness can be varied to create lighter and darker versions of a hue (see Fig. 13.25,

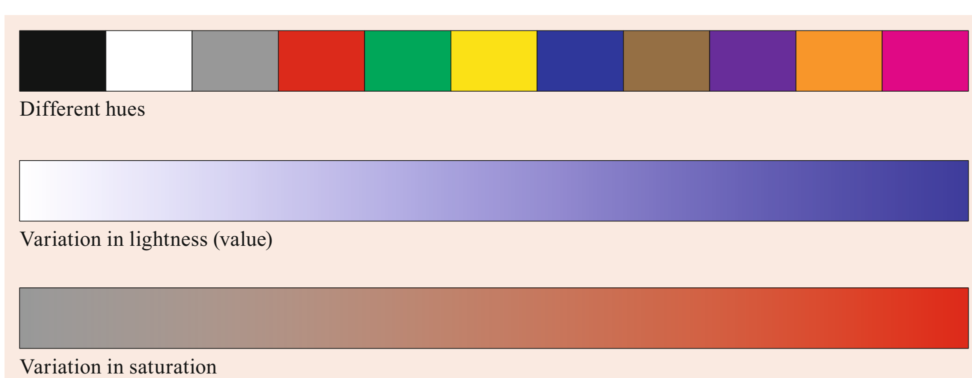


Fig. 13.25 The elements of color

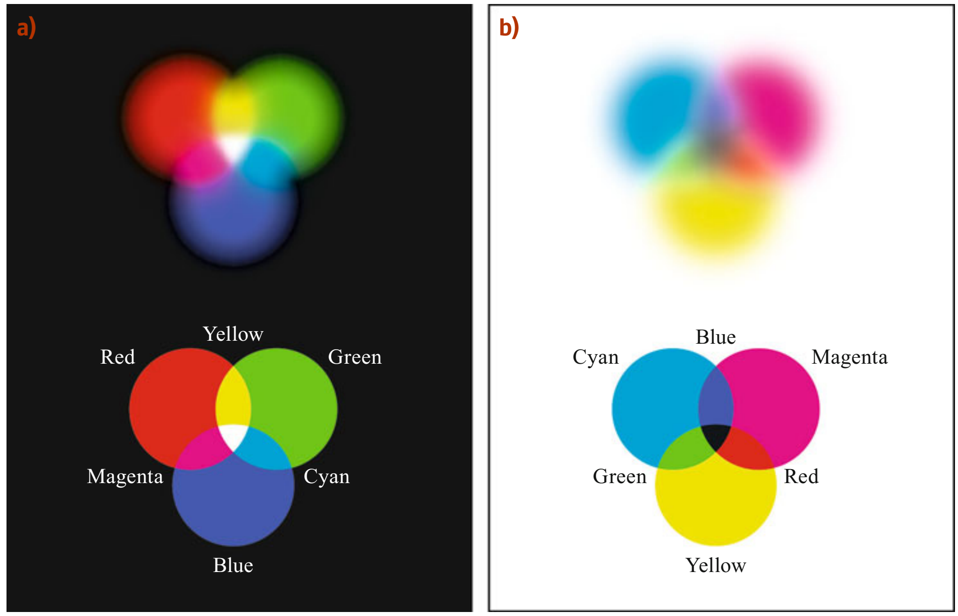


Fig. 13.26 The additive, RGB (**a**) and subtractive, CMY (**b**) color mixing systems

middle). Color saturation can change the colorfulness of a hue (Fig. 13.25, bottom). However, *lightness*, *darkness*, and *colorfulness* are relative concepts. If a map printed on paper is seen under a bright light, then the amount of reflected light is increased, and the lightness of all hues on the map will also be perceived to increase (although the relative differences between values of the hues will remain the same). This principle does not apply to a map displayed on a computer screen. In this case, the increased light of the hue is added to the light emitted by the monitor, resulting in a decrease in the relative differences in brightness in the screen image. This is one reason why problems arise when designing color for a print product on a computer and why it is best to work with color specifications rather than apparent color (an observer's perception of color, which varies depending on the light in the viewing environment).

13.5.2 Color Mixing

For cartography and other modes of graphic communication, color mixing is often based on two primary systems that relate to the technology used for production and presentation of the graphics (Fig. 13.26). The difference between the two systems lies in whether the hues are mixed as light sources or as ink pigments. For light sources, the RGB system is used. This is so named because the primary hues are red, green, and blue (RGB). For mixing ink pigments, the CMY system—based on the primary colors of cyan, magenta, and yellow (CMY)—is used.

The process of mixing colors using the RGB system is called additive color mixing. Red, green, and blue are referred to as the additive primaries in this color system. If three lights have red, green, and blue filters attached, they create beams of red, green, and blue light. Where all three beams overlap, the reflected light is perceived as white. Where only two beams overlap, the reflected light produces a mix of color. Blue and green mixed together form cyan, blue and red produce magenta, and red and green make yellow (Fig. 13.26a, bottom). By altering the intensity of each of the beams of light, different colors can be created, as in the interior space in Fig. 13.26a, top. Additive color mixing is used to display colors on computer screens in which extremely small phosphor dots arranged on a liquid crystal display (LCD) monitor emit red, green, and blue light in varying intensities. From a distance, the varying intensities reveal a sum of light and, consequentially, a color. The background color on a monitor is black, which represents no color emission.

Printing on white paper is achieved by using subtractive color mixing of three transparent inks—cyan, magenta, and yellow. These colors are referred to as the subtractive primaries or process colors. Cyan transmits green and blue light, magenta transmits red and blue, and yellow transmits red and green (Fig. 13.26b, bottom). For example, if yellow ink is printed on top of cyan ink on white paper, the yellow ink absorbs blue light but transmits red and green—the cyan ink then absorbs red light, so only the green light reaches the paper and is reflected back to the observer. Subtractive color mixing is used for digital printing with toners (like in laser printers) or liquid ink (for larger printers), and for offset printing in which the image is transferred, or offset, from a plate to a rubber blanket, and the image on the blanket is rolled onto a sheet of paper.

Because ink tones are not completely pure (in practice, they do not absorb or transmit 100% of the theoretical wavelengths of light), mixing the three subtractive primaries does not produce pure black. Hence, a black ink is usually used as a fourth printing color. Because black is normally printed first and the other colors are *keyed*, or registered, to it, it is assigned the acronym K. Thus, you will also see this referred to as the CMYK color mixing system.

13.5.3 Common Color Models

Colors can be defined using different color models, which are systems for specifying colors numerically according to their individual components. A color model is visualized as a three-dimensional space that contains combinations of the model's primary colors to produce all possible colors. Although a variety of color models has been developed, the RGB, CMY, and HSV models are primarily used in GIS.

The RGB color model specifies color based on the relative intensities of the red, green, and blue additive primaries on a computer monitor. The numerical values range from 0 to 255 for each of the three colors yielding 256^3 (16 777 216) possible colors. The primary colors can be represented on three axes of a cube, so that the cube becomes the color space in which any point can be given a numeric value for red, green, and blue that precisely specifies the color (Fig. 13.27). In this color model, black is defined as 0,0,0 and white as 255,255,255. Gray is scaled along the diagonal line between the two opposing corners between black and white. The secondary colors (cyan, magenta, and yellow) are situated at the remaining corners of the cube.

In the CMY color model, the CMY primary colors are analogous to the RGB primaries. The CMY color space can also be visualized as a cube with its primary colors at three corners, and its secondary colors (red, green, and blue) at the opposite corners (Fig. 13.28). Black and white are in opposite positions from their locations in the RGB color model,

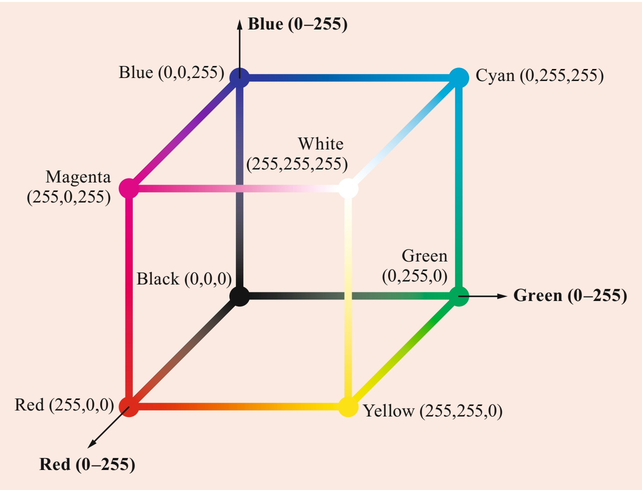


Fig. 13.27 The RGB color model

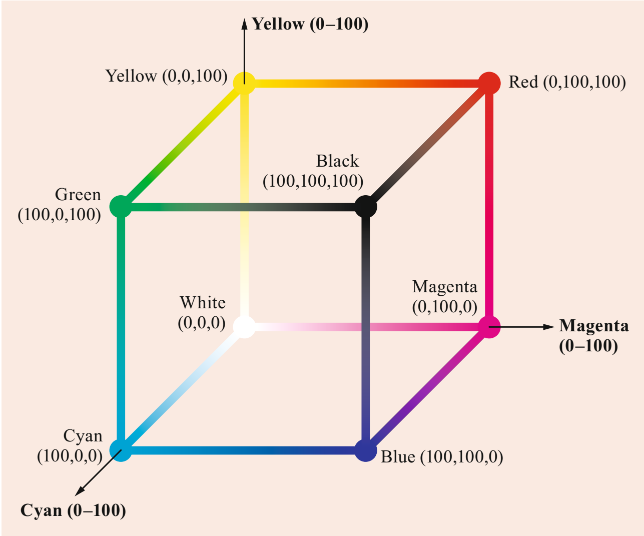


Fig. 13.28 The CMY color model

and gray is scaled between these two corners. Numeric values for the primary colors in this color model are expressed in percentages and range from 0 to 100 (although you will sometimes see the values ranging from 0 to 1 instead).

Another common color model used in GIS is the hue, saturation, and value (HSV) color model. The HSV color model can be visualized as a cone in a cylindrical coordinate system (Fig. 13.29). In this color model, the hues are arranged in the circle that represents the base of the cone inverted in Fig. 13.29 (for visualization purposes) so that the hue element of color can be expressed as an angle varying counterclockwise from red at zero. Saturation ranges from 0% (gray) at the middle to 100% at the face, or the curved surface of the cone's exterior, where the colors are fully saturated. Value ranges from 0% (black) at the apex of the cone to 100% at the base. White is located at the center of the base as the point with 100% value, 0% saturation, and no hue. Gray tones fall along the vertical axis from this point to the apex of the cone, where black is defined as 0% value, 0% saturation, and no hue.

The HSV color model is actually a rearrangement of the geometry of the RGB and CMY color models. It is more intuitive to many mapmakers because it allows them to work directly with the hue, saturation, and value (lightness) visual variables, as opposed to proportions or percentages of color determined by the technology used for production.

13.5.4 Design Guidelines for Color

When designing cartographic symbols, some relationship should exist among the color that is chosen, the data it is portraying, and the meaning to be communicated. For example, bare earth might best be symbolized with a brown hue because that is how dirt appears in the environment

13

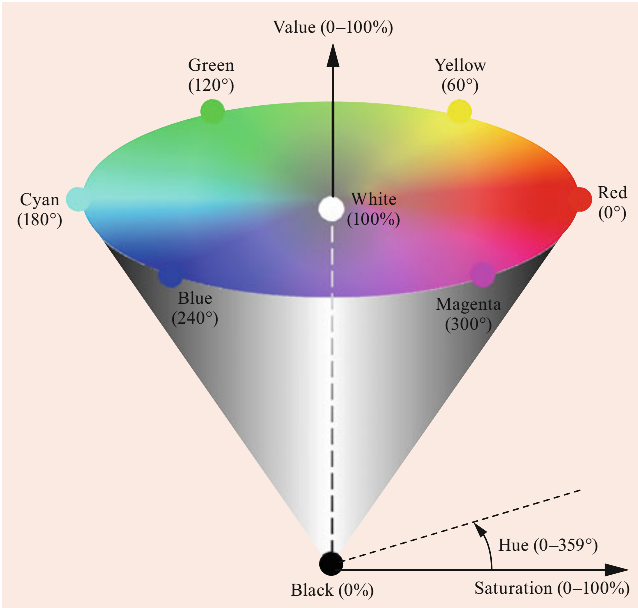


Fig. 13.29 The HSV color model

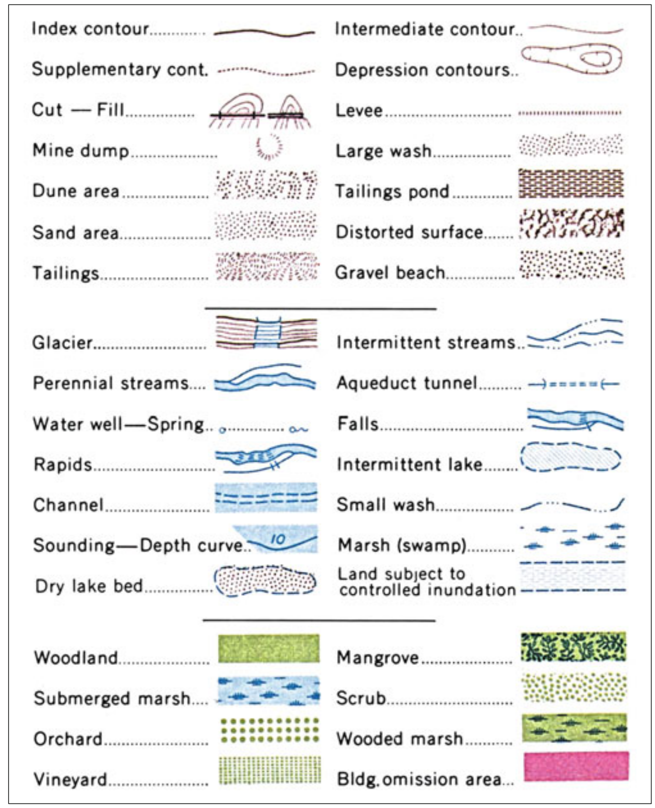


Fig. 13.30 Some of the symbols for the 1 : 24 000-scale USGS topographic map series (courtesy of USGS)

(Fig. 13.30). For some maps, the decisions about color and many other properties of symbols are dictated by specifications that override any creative decision-making on the mapmaker's part. For example, symbols on the 1 : 24 000-scale USGS topographic map series have been meticulously specified and the mapmaker is obligated to adhere to the relegated standards (Fig. 13.30).

Over time, a history of using certain colors for specific features has resulted in cartographic color conventions. Examples include blue for water features, green for forested areas, brown for contours, and black for political boundaries. It is important to keep in mind that color and other symbol conventions vary by locale and map type. For example, in one country's topographic map series, major roads may be shown with a line symbol that has black on the outside edges and a color, such as red, in the middle (this is an example of a cased line symbol), as with the first four symbols in Fig. 13.45. In another country, the interior of the symbol may be shown in a different color, for example, blue and pink in Great Britain (see the road examples in Fig. 13.21). It is also important to note that color conventions can conflict, as when two different symbols are used to show two different attributes of a feature. For example, an area may be in a certain jurisdiction that should be colored, for instance gray, and it may also be vegetated. If jurisdiction is the more important attribute for the map's purpose, its symbol will take precedence over the green fill that would conventionally used for vegetated areas. Nonetheless, color conventions allow mapmakers some decisiveness in their use of color.

When color conventions do not exist or maps are not designed to meet a specification, mapmakers can fall back on color guidelines developed through practice or research in cartography. A few are listed here as examples.

- When possible, colors that mimic the visual appearance of the features being mapped should be used, as with the green symbol for woodland in Fig. 13.30.
- To show associated features, color hues that are related in some fashion should be used. For example, different green colors could be used to show variety in the types of forested areas.
- The background of a map is often shown with a light, neutral or near-neutral color to ensure that it recedes visually; the foreground detail on a map is often symbolized using darker or more saturated colors.
- Symbols can be augmented with other visual variables, such as the submerged marsh and wooded marsh symbols in Fig. 13.30 in which the blue fill indicates submerged areas and the green fill indicates wooded areas.
- The emotive use of color should be avoided (for example, using highly saturated reds to symbolize disease).
- To draw attention to potentially ignored, small, or prominent features, use more saturated or darker colors than the surrounding areas.

Once the color use guidelines are known, they can be broken with a mindfulness toward achieving a specific goal or affective design (the look and feel of the map). For example, the use of darker colors for the background and lighter colors for the thematic information is a departure from the norm, so it can create a visual impression that is unique and, therefore, intriguing to map readers. The same holds for maps in black and white, since most maps today are made in color.

Color is extremely important in map design; therefore, understanding color definitions, perceptual qualities, and cognitive associations is helpful to mapmakers. For more on color in cartography, see [2] and [9].

In this section, much of the discussion focused on the use of color for reference maps. In the next section, the focus is on color use for thematic maps.

13.5.5 Color for Qualitative Thematic Maps

When considering color for thematic maps, cartographers often choose among cartographic color schemes, which are sets of colors to symbolize qualitative data categories or quantitative data classes on a map. Qualitative color schemes with varied hues should be used to represent different types of features (Fig. 13.31). For example, the land cover map in Fig. 13.19 has categories for water, developed land, forests, pastures, crops, and more. Each category should be visually distinct, without giving prominence to any one category or suggesting order in the categories. The simplest way to achieve this is to ensure that the different hues maintain a similar contrast with the background color of the map by controlling the value and saturation of each color. In general, this is not a problem for maps on which a single color or a limited number of colors are used for the background or maps with a limited set of classes or categories. However, this can be problematic on other types of maps, such as a thematic map overlaid on an aerial or satellite image in which the background has a lot of variation.

13.5.6 Color for Quantitative Thematic Maps

Color lightness, color saturation, or their combination should be used to represent order or numerical differences in quantitative attributes. In general, larger values or magnitudes should be shown with darker and/or brighter colors. Good examples of quantitative (also called sequential) color schemes are shown in Fig. 13.32.

The order of the quantitative classes shown on a map is also an important consideration in the selection of colors for a color scheme. On maps in which the areas between isolines are colored (Fig. 13.10), each class will always be portrayed next to its adjacent class (if the class is at the extreme of the range) or classes (if it is in the middle of the range). In contrast, on maps like choropleth maps (Fig. 13.5), any class has the potential to appear next to any other class or the same class. This can lead to the problem of simultaneous contrast in which two colors, side by side, influence each other and change the way we perceive each color. The effect is even more noticeable when complementary colors, such as red and green, yellow and purple, or orange and blue, are involved.

The number of classes used to portray the range of quantitative data is also important to consider, because it will impact how visually distinct each class is from the others. The smaller the number of classes, the larger the visual distinction, and vice versa. This gives rise to the conventional cartographic wisdom that choropleth maps should have no more than eight classes, although for some hues, like yellow, the number of classes should be even more limited. When too many classes are shown with a single color hue, map readers will find it difficult to determine to which class a feature on the map belongs, even when they study the legend. With isoline maps, the number of classes can be greater because of the forced adjacency of classes and their

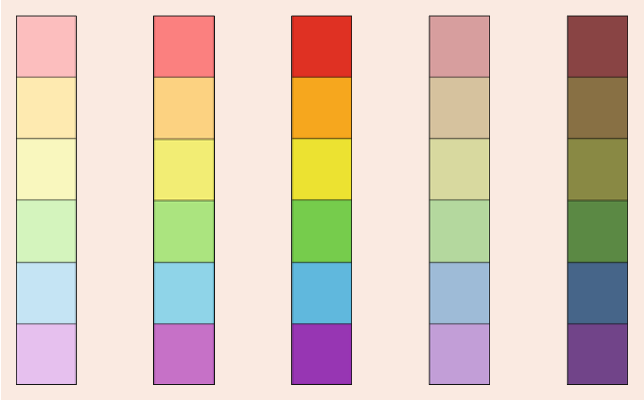


Fig. 13.31 Qualitative color schemes

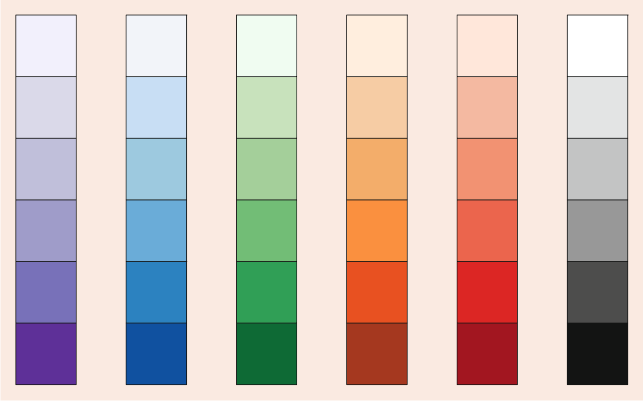


Fig. 13.32 Quantitative (sequential) color schemes

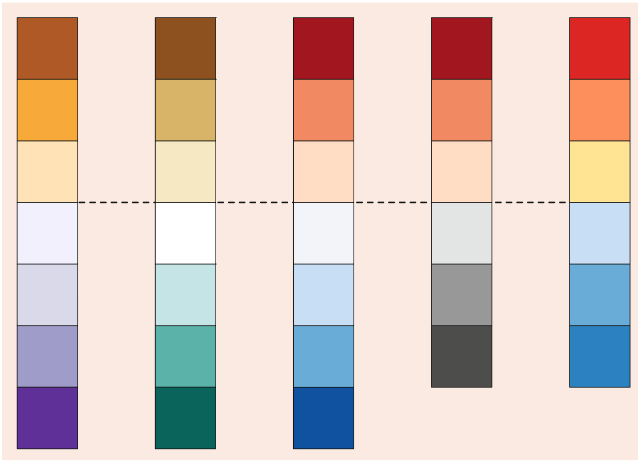


Fig. 13.33 Diverging color schemes

colors on the map as with the 11 classes in the legend in Fig. 13.10.

Figure 13.10 is shown with a diverging color scheme (sometimes called a bipolar color scheme) which is used for data that varies from a midpoint or critical value. A diverging color scheme portrays the midpoint or critical value with a neutral or near-neutral color, and values above and below that are shown with progressions of two different hues varying in lightness and/or saturation (Fig. 13.33). The midpoint or critical value can be either the single class in the middle or the break between the two different color progressions.

Choosing an appropriate color scheme is a challenge for many mapmakers, even seasoned professionals. A great resource is ColorBrewer 2.0 [10], an interactive web-based tool that presents color scheme options based on the mapmaker's requirements, such as the nature of data being mapped, number of classes, color and complexity of the background, usability for color blind readers, print friendliness, and ability to be photocopied. Like the examples in Figs. 13.31–13.33, ColorBrewer shows colors for symbolizing areal features; however, the same colors can also be applied to the design of point and line symbols.

13.6 Type

The great majority of maps make use of cartographic type to convey important or additional information about mapped features. Typical uses of type on a map include labeling places with their names ("Fort Scott" in Fig. 13.2) or numbers (101 the highway shield in Fig. 13.2), labeling features with values (as with the elevation values shown on the contours in Fig. 13.2), describing the environment ("Fort Point National Historical Site" in Fig. 13.2), locating a feature without using symbology ("Baker Beach" in Fig. 13.2), or indicating features of vague extent ("South Bay" in Fig. 13.2). Type is also used in titles, legends, text blocks, and other elements that accompany the map (see Fig. 13.5, for example). As with color, mapmakers benefit from basic knowledge about the properties of type and have, over time, developed guidelines on how to design and place type on maps.

13.6.1 Properties of Type

As in any profession, type designers have a specialized vocabulary to describe type and its properties (Fig. 13.34). Familiarization with this terminology makes it easier to communicate about typefaces and their properties and to recognize the underlying structure of various designs and the differences among them.

A typeface is a group of characters—letters, numbers, punctuation marks, mathematical symbols, and alternate or other characters—that share a common design or style. Helvetica, Arial, and Palatino are examples of typefaces. The different options available within a typeface make up a font family or type family—a set of characters that have the same basic qualities in their design, although their sizes, styles, and weights can vary. For cartographic purposes, we can divide typefaces into two main categories: serif and sans serif. A serif is a short line or stroke appended to or extending from the open ends of a character (Fig. 13.35). Sans serif (which literally means without line) fonts do not have these appendages.

A font is a set of characters within a typeface. A font is distinguished by its form (upright or italic), weight, width, ornamentation, and designer or foundry.

While the fonts within a family will vary based on their form, weight, and other characteristics, their essentially similar design is what ties them together. Thus, they are a family of fonts or a type family. Examples of font families for the Arial typeface include Arial, **Arial Black**, Arial Narrow, **Arial Rounded MT Bold**.

Many type families are available, at a minimum, in styles that include weight (such as ultralight, light, demibold, and bold or black) and form (upright, also known as roman, and italic, which is a uniquely designed version of a typeface that slants from left to right). Other families are much larger and contain options for additional variations, such as the width of the characters (for example, narrow, which is also sometimes called condensed, and expanded) and ornamentation (serif or sans serif). Size is another type property. This is generally expressed as a number of points (there are 72 points in an inch) or count of pixels (which is a digital expression) representing the height of the characters.

All these properties of type can be varied and give a font its unique appearance. Figure 13.34 illustrates a small set of

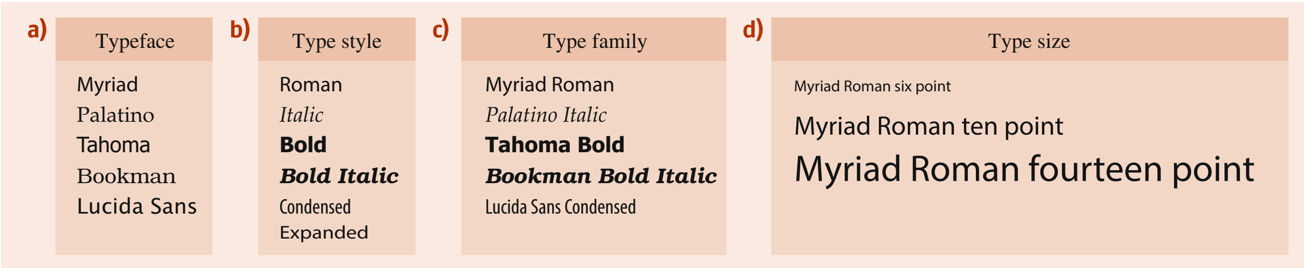


Fig. 13.34 Variations in type

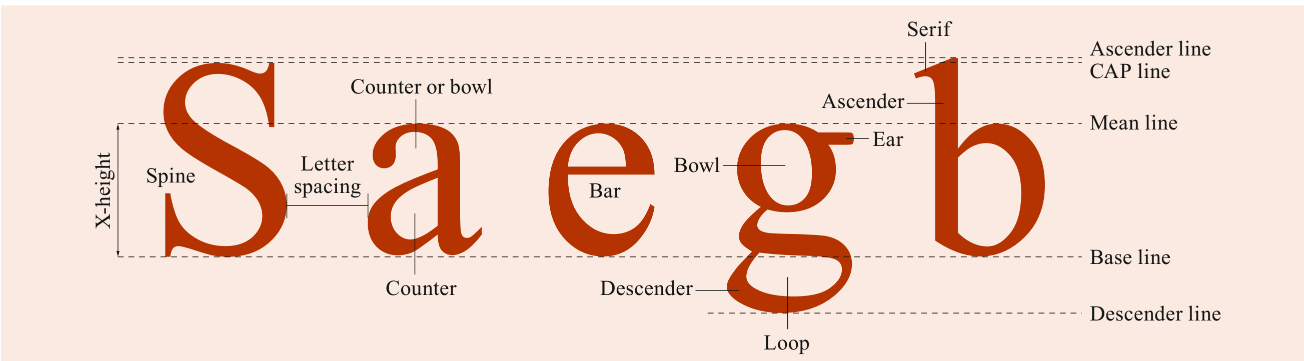


Fig. 13.35 The anatomy of a character (adapted from [11])

typefaces; some possible variations in style; and font or type families, which can then be varied in size.

A font can also include uppercase, mixed case (also known as title case), or lowercase characters. Some fonts are designed so that all characters are in one of these cases, but for most fonts, the user can set the capitalization of the characters. The user can also underline the characters.

There are thousands of typefaces, with new ones being developed all the time. So how can one typeface be distinguished from another? One important step is to train your eye to notice the details that set one font apart from another, as described above. Another is to examine the anatomy of the characters, as illustrated in Fig. 13.35.

For mapmakers, an important consideration of a character's anatomy is whether it has serifs, and whether to use a serif or sans serif font is one of the first decisions a cartographer will often make when choosing a font. Serif fonts are considered decorative and appear somewhat old-fashioned, but they have also been credited with increasing readability, especially for longer blocks of text, because they help the eye travel more quickly and easily across a line of type. For mapping purposes, labels are generally short and small, making a good case for the use of a sans serif font, which often has a more modern appearance. Some sans serif fonts are more legible than serif fonts at any size, especially if the sans serif fonts have open counters, a large x-height, and wide letterspacing (Fig. 13.35). These types of fonts are especially good for web maps, which are often displayed at a lower resolution on-screen than maps that are printed (especially professionally) on paper.

13.6.2 Design Guidelines for Type

A selection of general guidelines for the design of type used on a map is offered here. See Brewer's book [9] for a more complete explanation and illustrations.

- The number of font families used on a map should generally be no more than three.
- Short labels, as for contours, should be shown with a sans serif font.
- When there is a scarcity of space, a condensed font can be used to fit the label into the space available.
- A serif font should be used for large blocks of text.
- Mixed case should be used for most labels on a map.
- Larger, bold, uppercase, and/or underlined fonts can be used to label larger, more important, or prominent features.
- An italic font should be used to label water bodies and other features in the natural landscape; an upright (roman) font should be used for cultural features.
- Color can be used to distinguish the labels for different types of features. For example, a black font can be used to

label cultural features, brown to label terrain features, and blue to label hydrographic features.

Guidelines such as these offer mapmakers a good foundation from which to start making type design decisions. Departures from these norms can lead to interesting design alternatives that may improve a map. However, the legibility of labels is paramount (see Sect. 13.8.1 for more on legibility). Type overlap (or overlap, for short) occurs when text is overlaid on a symbol or other text. This not only affects legibility but also looks sloppy and should be avoided if at all possible.

13.6.3 Placement Guidelines for Type

Designing type on maps requires consideration not only of the appearance of the type but also its placement. Placing type *by hand* (positioning each label individually, even on a computer) is a slow and painstaking task. Most GIS systems have the capability to generate labels from attribute values and place the labels relative to their associated features. Labeling engines allow the user to specify rules for both type appearance and placement. Advanced labeling engines allow mapmakers to define a greater number of parameters, particularly for placement, as well as assign label and feature weights. These weights are used to specify the relative importance of labels and features to resolve conflicts when there is a situation involving overlap.

There is a tradeoff between simple but fast labeling engines, which may result in labels in dense areas being placed on top of one another or on underlying features, and intelligent labeling engines, which iterate solutions to find the best places to position labels for minimum conflict (that is, overlap). Advanced labeling engines can even change properties of the labels to fit the geographic context (for example, expanding the letter and word spacing to fill a large area or condensing the spacing or using an abbreviation to fit a label into a tight space). Advanced label placement requires contextual analysis and conflict resolution, which can be computationally intensive tasks.

When placing labels manually or setting up rules for automated label placement, cartographers often rely on best practices that they have developed practically and tested scientifically over time. Their first source of reference for label placement is often Imhof's article [12]. Imhof starts by advising that placement guidelines are always dictated by map scale. For example, on a large-scale map, the label for an areal feature may fit within the area, but on a small-scale map, the label should be placed as though it relates to a point or linear feature (Fig. 13.36).

Imhof's treatise on label placement is quite extensive, and it includes labeling the following: point, line, and area features (an example for point features is shown in Fig. 13.37); points along linear features, shorelines, or coastlines; mountain features; and features on polar maps. Imhof also offers guidance on lettering direction, label spacing, label overlap, type combined with other map contents, and the total impression of lettering. Instead of having to remember all these details, mapmakers can take comfort in knowing that many of Imhof's guidelines have been encoded in GIS labeling engines.

Fig. 13.36 Label placement guidelines vary with map scale: Large scale (**a**) and small scale (**b**) examples are shown for Puerto Rico; the Columbia River, which is the border between Washington state to the north and Oregon to the south in the United States; and El Salvador (data courtesy of Esri)

Although mapmakers attempt to avoid placing type over other symbols or labels on the map so that the type is not difficult to read, this is sometimes impossible, especially when type is placed over features to label them (for example, contour labels over contour lines). When it is not possible to

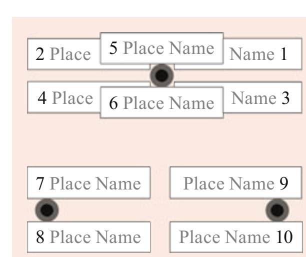


Fig. 13.37 Preferred label positions for point features (adapted from [12])

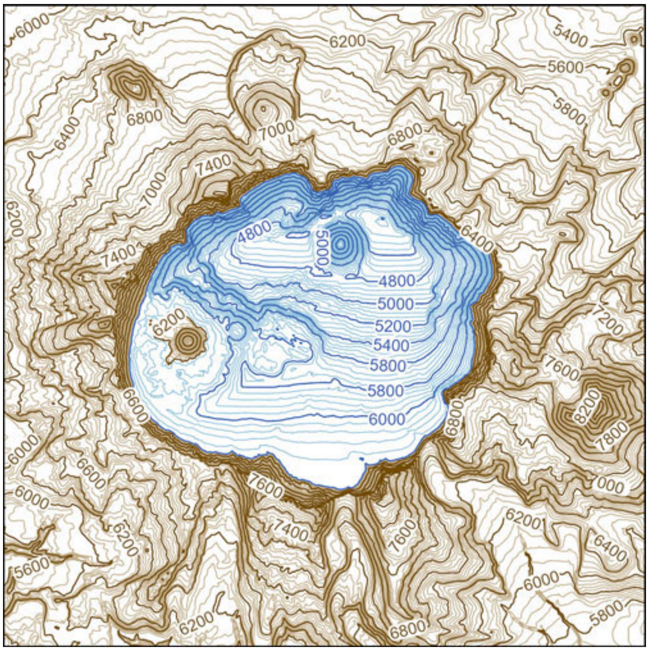


Fig. 13.38 Contours and bathymetric contours and their labels for Crater Lake National Park in southern Oregon, USA (data courtesy of Esri)

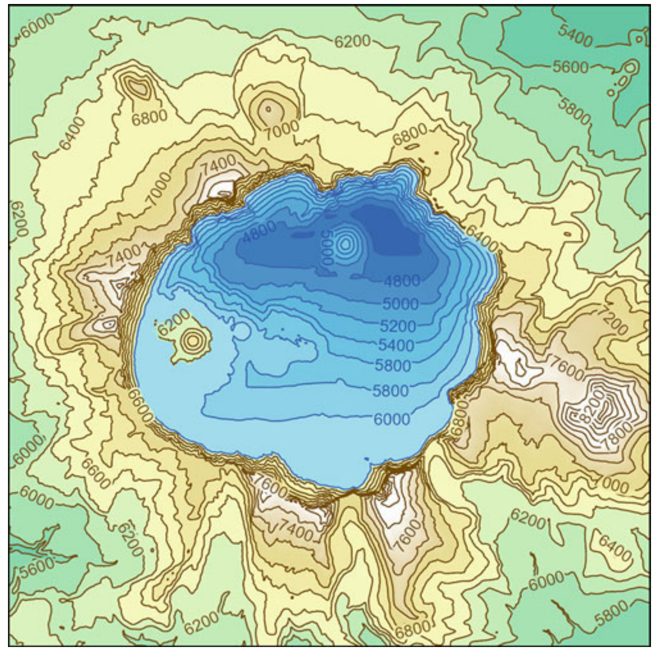


Fig. 13.39 Layer tints for the terrain and bathymetry (data courtesy of Esri)

avoid such overlap, cartographers rely on techniques to mitigate the impact of overprinting. A common technique is to use a halo—a buffer around the label filled with the color of the background, such as the contour labels with white halos in Fig. 13.38. Another technique is to knock out, or erase, only the conflicting areas of the underlying symbols or labels, as with the contour and bathymetric contour labels in Fig. 13.39. This is often a better solution when the background is multicolored.

13.7 Relief Portrayal

Relief portrayal is used on maps to show the topography (elevation) and form of the land surfaces, as well as the bathymetry (depth) of water bodies (Figs. 13.38 and 13.39). Because GIS has made it easy to display relief on maps; however, if the relief is not directly relevant to the subject of the map, as is often the case for thematic maps, then it should not be portrayed, as this could confound the message of the map and compromise the user's ability to clearly see the most important features. A definitive work on the theory and methods of relief portrayal in cartography was written by Imhof [13]. Many of his methods have been encoded in GIS software, resulting in the ability to portray the land's form in many interesting and useful ways, often with relative ease. GIS can be used to portray absolute relief by showing precise elevation (distance above sea level) and depth (distance below sea level) information and relative relief to give a general impression of relative heights and depths. Both absolute and relative relief can be shown on maps, and often they are shown together, as described below.

13.7.1 Absolute Relief Portrayal

One of the easiest ways to show relief is to label the elevation of surveyed locations on land with spot heights, sometimes improperly called benchmarks (see the small black labeled circles in Fig. 13.46), and labeling depth under water with soundings (see the gray labels in Fig. 13.4). Maps with these types of labels are often augmented with contours (isolines of equal elevation) or, for water areas, bathymetric contours (also called isobaths or depth curves). For terrain, the contours are traditionally shown as brown lines at a specified contour interval (the difference in value between contours), as illustrated in Fig. 13.38. Every fourth or fifth line (depending on the contour interval) is labeled and shown with a thicker line (these are called index contours), while the remaining intermediate contours are thinner and sometimes lighter in color (Fig. 13.30).

Despite the simplicity of the concept of contours, many map readers have difficulty interpreting relief based on this mapping method. Mapmakers will, therefore, sometimes color between the lines, creating what cartographers call a layer tint (Fig. 13.39)—this is called a hypsometric tint for land elevation and a bathymetric tint for water depth. A layer tint aids in interpretation if the colors are chosen appropriately. The goal is to select colors that mimic the land cover

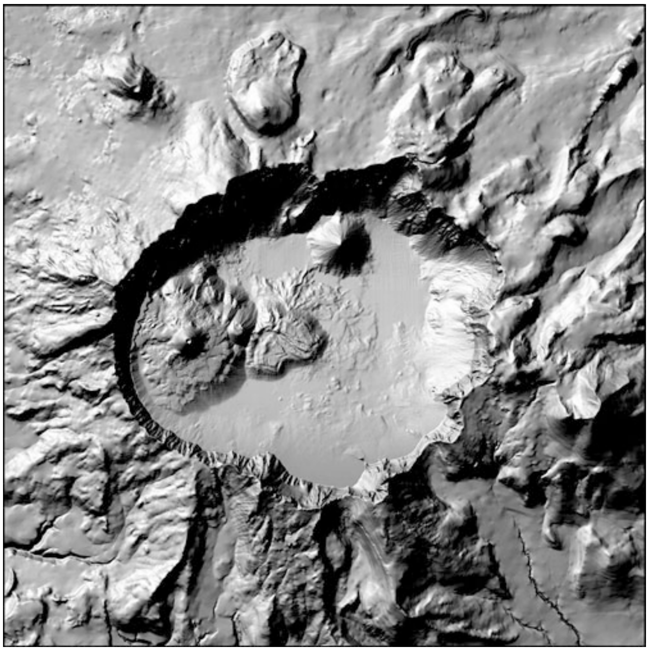


Fig. 13.40 A default hillshade (data courtesy of Esri)

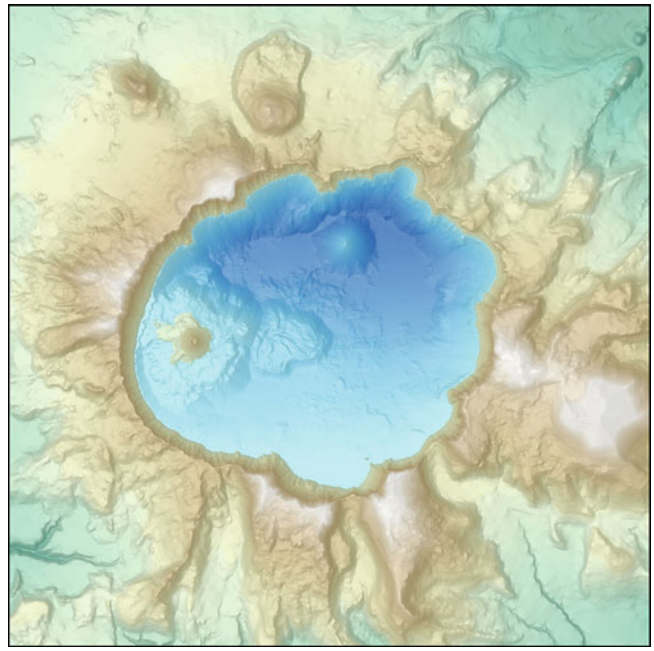


Fig. 13.41 A multidirectional hillshade overlaid with a hypsometric tint (data courtesy of Esri)

or give an impression of the depth of the water. For example, in the Crater Lake area shown in Fig. 13.39, lowlands and valleys are shown in green; higher, scrub-covered elevations are yellow-brown; treeless mountaintops are brown; and the snowcapped peaks are white. The bathymetry of the lake is shown in blue, with deeper depths having darker shades.

13.7.2 Relative Relief Portrayal

Elevation data is often available in a raster format called a digital elevation model (DEM) in which each pixel or cell has an elevation or depth value. DEMs can be used to generate a variety of relief representations, of which the most common is a hillshade. Hillshading (also called shaded relief and relief shading) simulates the light and shadows that would be cast on the surface from a source of illumination, such as the sun. In Fig. 13.40, the illumination is from a northwest source that, by default, is at 45° above the horizon and 315° from north.

A relief portayal method that is currently popular is called multidirectional hillshading, which involves illuminating the surface from light sources in multiple directions, although the strongest light still comes from the northwest, as in Fig. 13.41. Hillshades are often overlaid with a layer tint, which can also be generated from a DEM. While the colors are again selected to simulate the land cover or the depth of the water, the visual effect is one of continuous change rather than abrupt breaks at contours (Fig. 13.41). This eliminates the need for contours and labels, freeing the background to support overlaid features

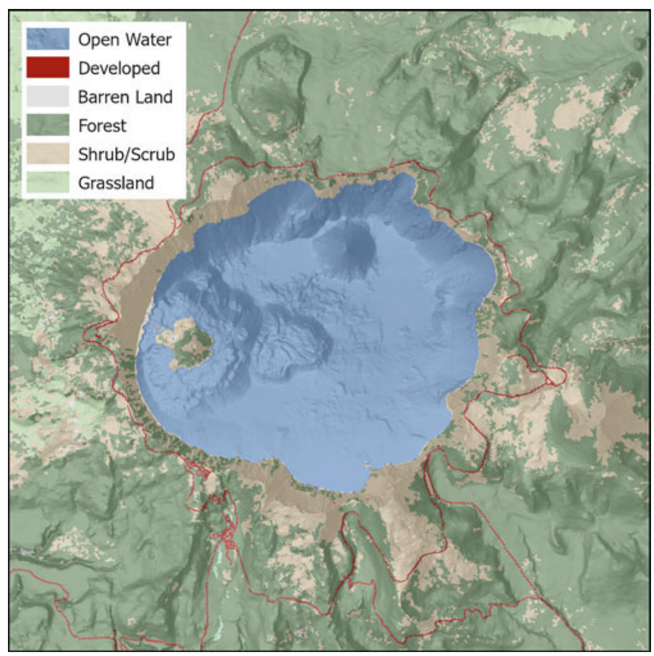


Fig. 13.42 A multidirectional hillshade overlaid with National Land Cover Data (NLCD) (data courtesy of Multi-Resolution Land Characteristics Consortium and Esri)

and their labels. However, the result is a relative representation of relief, rather than an absolute one.

For terrain surfaces, a hillshade can be a subtle graphic (using a range of light gray tints), so other layers, aside from hypsometric tints, can also be displayed over the surface. It is often best to overlay only those themes that have some rela-

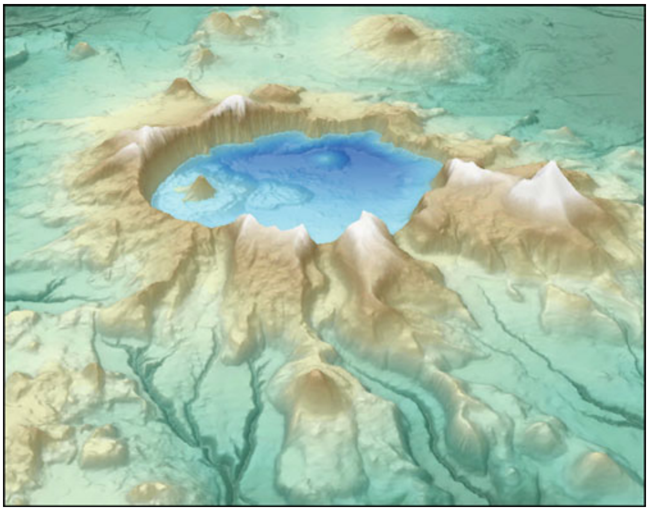


Fig. 13.43 An oblique view of the terrain with 3x exaggeration (data courtesy of Esri)

tion to terrain, such as geology or land cover (Fig. 13.42), but you will often find other non terrain-related layers displayed over hillshaded surfaces. It is also common to see an aerial or satellite image draped over the surface and sometimes even a scanned image of a printed map.

Another method of relief portrayal is to switch from a planimetric view looking down on the landscape from directly above to an oblique view looking across the surface from an angle. Sometimes, vertical exaggeration is added to the DEM values to make the view more dramatic. The terrain in Fig. 13.43 is exaggerated by a factor of 3, such that the elevation values in the DEM are multiplied by 3. For comparison purposes, the bathymetric data has not been exaggerated. It is important to remember when looking at these types of portrayals that the variations in relief on the ground are not as great as they appear on the map.

Three-dimensional cultural features, such as buildings, and natural features, such as trees, as well as other symbols and text can also be superimposed on the oblique view. For maps of this type, it is useful to provide the user with the interactivity to walk or fly through the view to inspect details and see obscured parts of the map.

13.8 Map Design

We have seen that symbolizing and labeling mapped features are major tasks in the mapmaking process. Another important step is to compose the map as a complete, clear, and compelling graphic communication product. When designing the map (or, at times, set of maps) and laying out the contents on the page or screen, cartographers employ many principles of basic graphic design. Knowing these design principles helps mapmakers to produce better maps and map readers to gain a better understanding of how maps are presented.

13.8.1 Legibility

For all symbols and type on the map, mapmakers must carefully consider legibility, which is the ability to be seen and recognized. The legibility of a symbol or label is related to its size and its contrast with the background and other symbols and labels (Sect. 13.8.2). Symbol and type size guidelines are offered in Tables 13.2 and 13.3. Because type characters tend to be more complex than simpler map symbols, the sizes for the type are slightly larger than for symbols at any viewing distance. For complex map symbols, it is safer to use the type size guidelines than the symbol size guidelines to ensure legibility.

A useful online resource to explore the design, arrangement, and legibility of symbols is the Map Symbol Brewer by Schnabel [14].

13.8.2 Visual Contrast

Visual contrast relates to the visual distinction between a symbol or label and what it overlays, such as the map background. Difficulties arise when the overlaid objects are too similar in appearance (for example, red text on an orange background), or when too many things are overlaid such that there is no space to see between objects to the background.

Table 13.2 Symbol size guidelines (adapted from [2])

Viewing distance (ft; ft = 0.3048 m)	Computer screen (pt)	Print maps (pt)	Computer screen (in.)	Print maps (in.)	Computer screen (mm)	Print maps (mm)
1.5	6	4	0.08	0.06	2.0	1.5
2	8	6	0.10	0.08	2.7	2.0
3	11	8	0.16	0.12	4.0	2.9
4	15	11	0.21	0.15	5.3	3.9
5	19	14	0.26	0.19	6.6	4.9
10	38	28	0.52	0.38	13.3	9.8
20	75	55	1.06	0.77	26.6	19.5
30	113	83	1.57	1.15	39.9	29.3
50	188	138	2.62	1.92	66.5	48.8
100	377	276	5.24	3.84	133.0	97.5

Table 13.3 Type size guidelines (adapted from [2])

Viewing distance (ft; ft = 0.3048 m)	Computer screen (pt)	Print maps (pt)	Computer screen (in.)	Print maps (in.)	Computer screen (mm)	Print maps (mm)
1.5	8	6	0.11	0.08	2.8	2.1
2	11	8	0.15	0.11	3.8	2.8
3	16	12	0.22	0.17	5.6	4.3
4	21	16	0.30	0.22	7.5	5.7
5	27	20	0.37	0.28	9.4	7.1
10	53	40	0.74	0.56	18.8	14.2
20	107	80	1.48	1.12	37.6	28.4
30	160	121	2.22	1.68	56.4	42.6
50	266	201	3.70	2.79	94.0	70.9
100	533	502	7.40	5.59	188.0	141.9

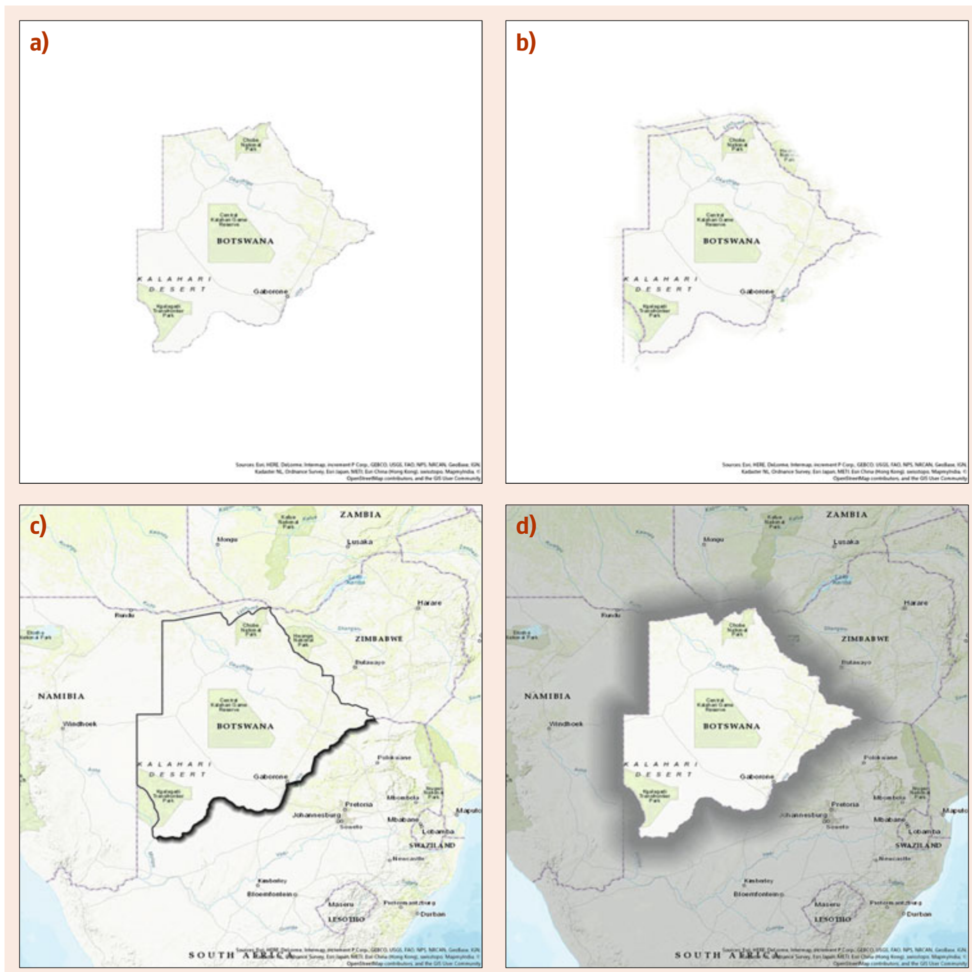


Fig. 13.44 Methods that can be used to promote figure-ground organization include a closed form (**a**), a vignette (**b**), a drop shadow (**c**), and illumination (**d**) (data courtesy of Esri)

Thus, maintaining visual contrast becomes more difficult as objects are added to the map. The most important design decision to consider in promoting visual contrast is choosing the colors of an object and its background. Map scale, function, and specifications (as illustrated in the 1 : 24 000 scale topographic map symbols in Fig. 13.30) will also impact the mapmaker's ability to maintain good visual contrast throughout the map.

The methods used to promote visual contrast also relate to the conditions of map use. For example, if the map is to be used in low-light conditions, the design will necessarily change. This is apparent on mobile maps (such as those displayed on smart devices) that automatically change from dark objects on a light background in daylight conditions to light objects on a dark background in low-light environments.

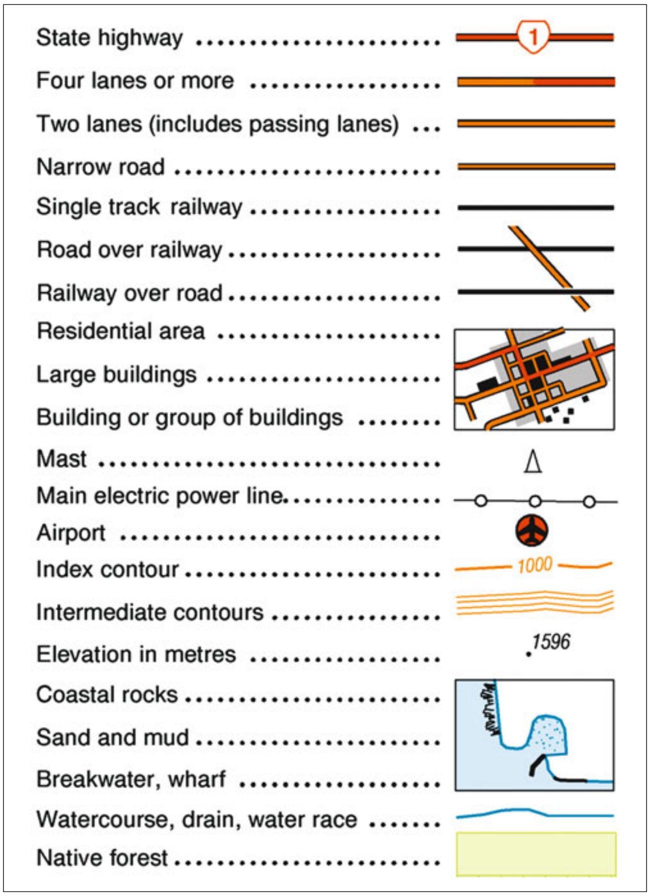


Fig. 13.45 Legend for the map in Fig. 13.46 (courtesy of Land Information New Zealand (LINZ))

13.8.3 Figure-Ground Organization

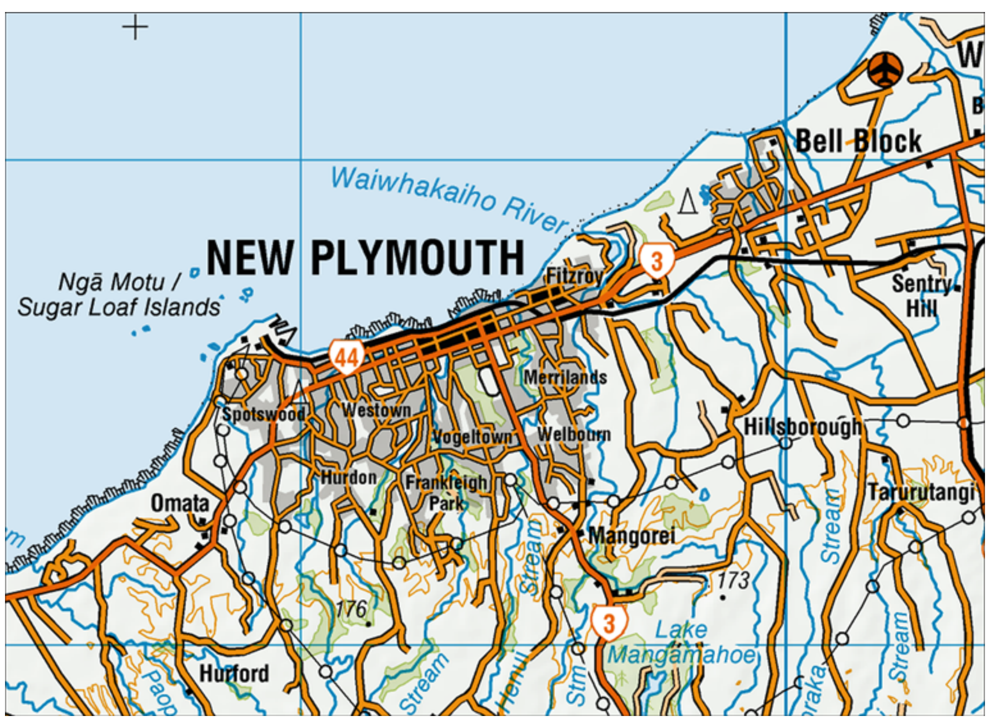
A graphic design concept that is similar to visual contrast is figure-ground organization, which is the visual separation of the map into the mapped area (the figure) that stands out against an amorphous background. Examples of cartographic methods to promote figure-ground organization are illustrated in Fig. 13.44. These include use of a closed form (a polygon that represents the entire mapped area), a vignette (a buffer around the perimeter of the figure that fades into the background), a drop shadow (a thin vignette along the perimeter of the figure offset in only one direction—generally, the southeast—giving the impression that the figure is raised above the ground), and illumination (a vignette that fades into a background that is darkened thus "illuminating" the figure). Other methods also exist, and many of the methods can be combined.

13.8.4 Visual Hierarchy

Visual hierarchy relates to the internal, graphic structure of the content on the map. It is used to communicate the relative importance of mapped features through a visual impression that categories of features are prioritized or ordered. Correctly applied, visual hierarchy reflects an appropriate intellectual order by graphically emphasizing the most important map details and de-emphasizing less important features.

The categories involved in the hierarchy and their order are dictated by the map's use. For reference maps and charts,

Fig. 13.46 A topographic map with good visual hierarchy (courtesy of LINZ)



categories of map features and their labels should be organized according to a hierarchy that relates to how the features are organized structurally in the physical landscape. For example, using the legend in Fig. 13.45, it is apparent in the 1 : 250 000-scale topographic map in Fig. 13.46 that terrain, shown with contours, is at the lowest level, followed by land cover (vegetation and built-up areas); hydrography (rivers, streams, and water bodies); transportation (roads and railroads); power lines; and, at the highest visual and conceptual level, place names and other labels. For thematic maps, all reference data may recede to the background, while the thematic information takes visual prominence on the top layer (see, for example, Fig. 13.11).

13.9 Layout

Map products often require design and compilation of more than the main, and often, single map, which is primarily what has been discussed to this point in this chapter. Other content contributes to the map's message by helping to explain what is on the map or providing information that cannot be portrayed on the map. What makes up this ancillary content is often referred to as map elements (or sometimes map marginalia, although it will not always appear in the margins). When the map and its elements are organized and displayed together, the concept of layout comes into play.

Layout is the arrangement of type, graphics, and space on the page (or in the case of web and digital maps, the computer screen). Layout can be used to create a story, provide a voice, promote engagement, and impart a style, all relative to the map's intended message and use. For expediency, in this chapter, we use the term *page* with the understanding that the discussion also applies to a computer screen.

Layout starts with the page—a blank space with visible and invisible properties. Other elements are added to that space. Additional graphic design concepts arise when considering not only the map and its internal content but also the external content (the map elements) shown with the map. Three of these concepts—proportion, balance, and harmony—are discussed later in this section.

13.9.1 Map Elements

Common map elements and some of their general design guidelines are presented below.

- The title and subtitle—an explanation of what is shown on the map, including the subject or theme, location, date, and other pertinent information (number 10 in Fig. 13.47).
- The legend—an explanation of the symbols and labels found on the map, usually shown in graphic form (see, for example, Fig. 13.5), although sometimes indicated as a word phrase (such as "One dot equals 5 people per square mile" in Fig. 13.14). The legend should only include symbols shown on the map or set of maps, and all symbols in the legend should appear exactly as they do on the map (the same size, shape, orientation, level of transparency, etc.).
- The grid or graticule—an array of north-south, east-west lines based on latitude and longitude (the graticule is shown with the blue lines in Fig. 13.46); a grid coordinate system, such as universal transverse Mercator (UTM); or a grid cell coordinate system, such as the Military Grid Refence System (MGRS). (See [2] for a complete explanation and illustrations.)
- The map scale indicator—an indication of the map scale in one or more of the formats described in Sect. 13.3.1 (numbers 13 and 14 in Fig. 13.47). For small-scale maps, scale will often vary across the map, so an indication of map scale will only be true at a point or along a line or lines—as such, map scale indicators are often not used on these types of maps, and the graticule can be used instead. Map scale indicators are not pertinent to oblique views of an area.
- The orientation indicator—an indication of the orientation of the map, often shown as a north arrow, although a compass rose is commonly used on charts (the letter c in Fig. 13.47). An orientation indicator is essential if the map is not in normal orientation (with north at the top of the page). An orientation indicator should not be used on maps in which the direction to north varies (as shown by the light gray lines indicating the graticule on the map in Fig. 13.1). In these cases, the graticule can be used instead.
- A locator map or location diagram—a small map or diagram (sometimes shown as an array of rectangles or trapezoids) that indicates the extent of area shown on the main map and sometimes the extent of adjacent maps.
- An inset map—a large-scale map that shows an enlargement of an area with dense symbology and/or labeling on the main map that would be difficult to see at the main map scale (number 18 in Fig. 13.47). Multiple inset maps may be used if there are multiple congested areas on the main map.
- Projection and datum information—text indicating the map projection and, especially for maps that are used to determine distance and direction, the datum, which is the reference system against which position and elevation and depth measurements are made (number 3 in Fig. 13.47).
- Credits and publication notes—text that indicates the author, source(s) of data (number 17 in Fig. 13.47), date of edition or publication (number 6 in Fig. 13.47), publication (number 4 in Fig. 13.47), and other pertinent information to authenticate the map.
- Neatline—an indicator of the map extent. The neatline may be simple (see the right and bottom sides of the map in Fig. 13.47), or decorative. It can also be anno-

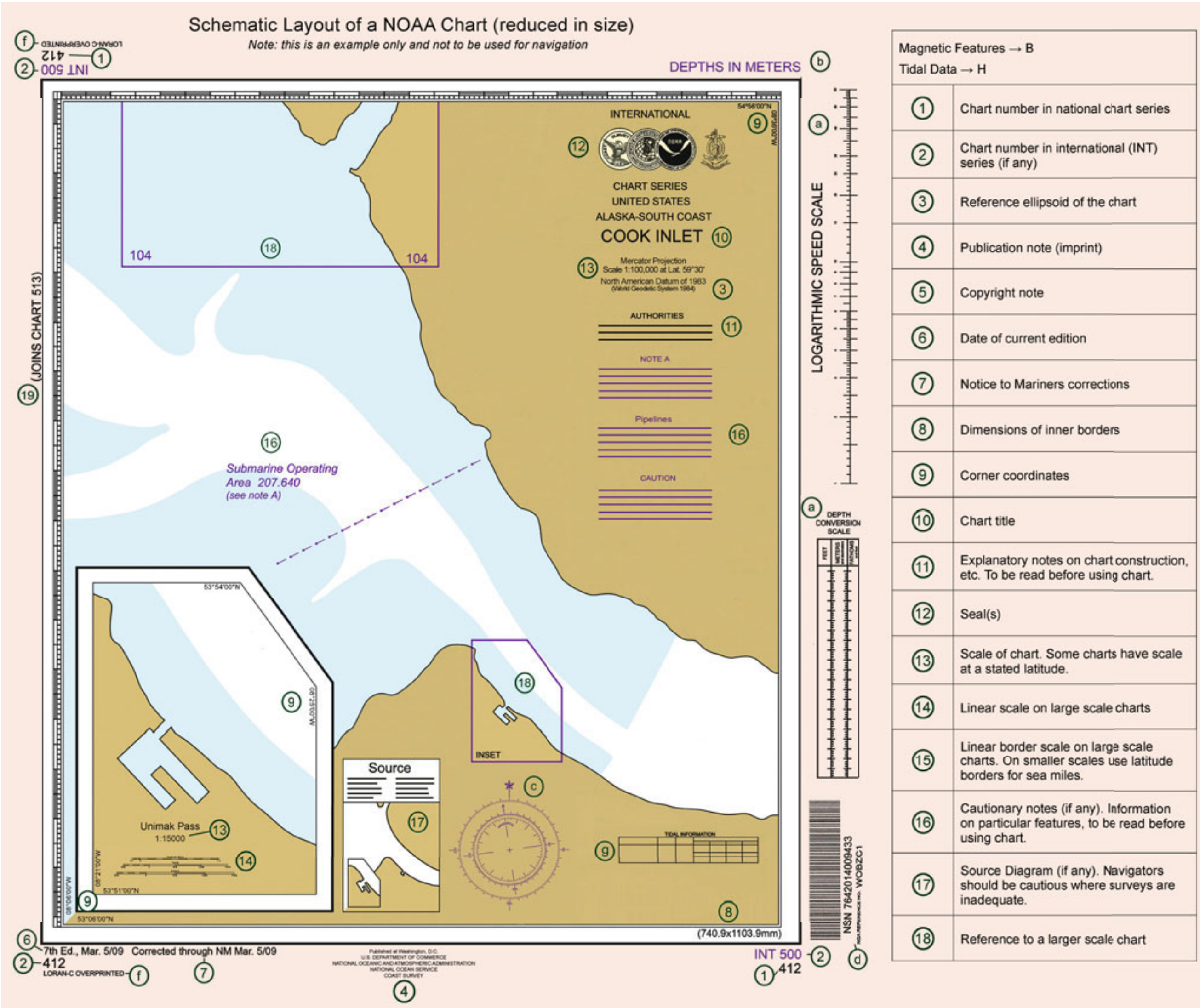


Fig. 13.47 The page layout for a NOAA nautical chart showing many of the common map elements (courtesy of and adapted from NOAA, https://nauticalcharts.noaa.gov/publications/docs/us-chart-1/ChartNo1.pdf)

tated and graduated to show the grid coordinate system or systems used on the map; this is common on charts and topographic maps. On most nautical charts, longitude is graduated along the top or bottom of the chart, and latitude is graduated along a side (see the top and left sides of the map in Fig. 13.47). The subdivisions are in degrees, minutes, half-minutes, and tenths of minutes—and, on older charts, in seconds. It is important to remember that only the vertical, latitude scale (but not the horizontal longitude scale) on a Mercator chart can be used to measure distance because a minute of latitude equals 1 nautical mile anywhere in the world, while the distance between minutes of longitude decreases between the equator and the poles.

- Text blocks, graphs, photographs, and other ancillary elements—additional content that helps explain the map and its intended message (for example, numbers 11 and 16 in Fig. 13.47).

Not all maps need all these elements, but some maps—such as topographic maps and charts—use many of them. Additional map elements may also be required, depending on the map type and intended use, as shown in Fig. 13.47.

With increased use of the web to share maps, many or all of the map elements may be compiled, along with the map (or several maps), in what is called a story map. A story map is a sequence of maps that narrate a story. It is often augmented with text, photos, and/or video, and the web application (or app) provides functionality, such as panning and zooming, pop-ups, a magnifying glass, a swipe tool, and/or a time slider—all of which help readers better understand the story.

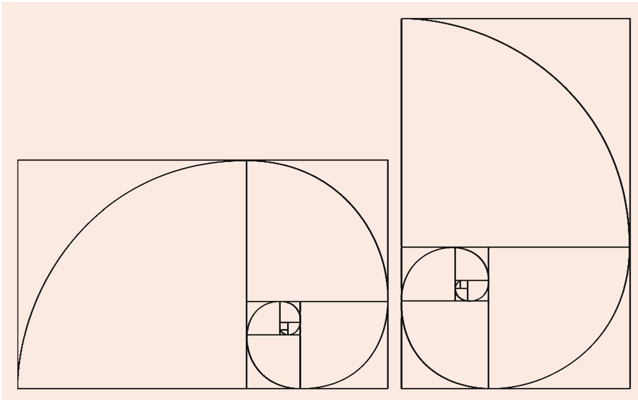


Fig. 13.48 The geometry of the golden ratio (3 : 5)

13.9.2 Proportion

Proportion is the relationship between one part or element of the map and another or the whole map. Proportion relates to both the map elements (for example, the legend) and parts of the elements (such as the legend boxes). Good proportion is achieved when a correct or desirable relationship exists between the elements with respect to size, shape, scale, color, quantity, degree, visual weight, or setting. Design concepts that relate to proportion include symmetry (one side being a reflection of the other), balance (equivalence in the visual weight of the elements), and harmony (elements appearing to belong together in size and distribution).

Symmetry is often not something that cartographers can control on their maps or, sometimes, even their layouts. The geography of the features in the mapped area may or may not be symmetrical (notice how Crater Lake in Fig. 13.38 is more symmetrical than California in Fig. 13.6). There is nothing to be done about this because that is simply the geography of the area being mapped. Also, notice that the map elements in Fig. 13.47 are not symmetrical, partly because there is a need for only one title, one legend, one scale bar, one compass rose, and so on. Sometimes symmetry can be achieved using groups of elements rather than individual elements. Balance and harmony are easier to achieve in a layout and are discussed further in Sects. 13.9.3 and 13.9.4.

When making decisions about where to place map elements in a layout, cartographers often rely on the principle of the golden ratio. This is a mathematical ratio found in both human design and nature that can, when used properly, help create aesthetically pleasing compositions in map design. The golden ratio (which is known by many alternate terms, such as the golden section, golden mean, divine proportion, or the Greek letter *Phi*) is approximately equal to 1.618, such that if one side of a rectangle is 1 unit, the other is 1.618 units. This ratio (which can also be roughly expressed as 3 : 5) is considered to produce images that are pleasing to the eye. Figure 13.48 illustrates this design concept.

The golden ratio relates to both the space on the page and the space used for the elements placed within those areas. For example, a legend designed using a 3 : 5 ratio will be more visually pleasing than one based on a 1 : 1 ratio. Similarly, if the legend itself contains rectangles to show areal feature symbols, the rectangles can also be designed using a 3 : 5 ratio to promote an aesthetically pleasing design. Notice how the design of the page and many of the elements in Fig. 13.47 employ the principle of the golden ratio.

13.9.3 Balance

Balance relates to the arrangement of the map and its elements on the page. The objective is to create visual equilibrium rather than the appearance of being off-balance in any direction. To achieve pleasing and effective visual balance, the page contents should be positioned around the visual center of the page, which is a point just above the geometric center of the page. This is the point on which the eye first focuses, and it serves as the fulcrum, or balancing point, for the page.

The impression of visual balance is controlled by the size, visual weight, and location of the symbols and labels on the map and the map elements on the page. For example, one large map element may tip the layout out of balance—to counter this, several smaller map elements can be grouped together and balanced against the larger element.

Poor visual balance leads the reader to see individual components as competing for space or to see an abundance of either congested or empty space. However, it is not always desirable to fill all the empty space in a layout. White space helps to give the impression of breathing room and create a sense of freedom for the eye to choose what to focus on. On the other hand, large unfilled spaces can throw the layout out of balance and give an impression of lack of attention to basic design. This can be mitigated in a number of ways, such as using a faded background (Fig. 13.44d) or showing the graticule outside the mapped area (Fig. 13.46).

13.9.4 Harmony

When a layout has harmony, the individual elements on the page present a meaningful whole. Harmony is achieved through the arrangement of elements to create a pleasing image in which the map and its elements complement each other and work together visually. As each element is added, its effect on the layout must be assessed, and use of the remaining space must be reevaluated. The layout may need to be reorganized several times to achieve harmony.

Harmony can be enhanced through the alignment and distribution of elements in the layout. Alignment is the arrangement—generally in an orthogonal direction relative to the page—in a straight line or in an appropriate relative position. Note, for example, that the elements at the right side of the map in Fig. 13.5 are aligned with the right side of the page, and the elements along the bottom are aligned with the bottom of the page. Proper alignment creates an ordered appearance by ensuring that the elements have an obvious and inherent relationship with each other. Proper alignment of elements eliminates the apparent messiness that can occur when elements are placed seemingly randomly. In most GIS systems, elements can be easily aligned using alignment tools or guides (usually dotted lines that appear in the layout view). Intelligent guides can be snapped to when elements are repositioned and provide an indication of when the elements are properly aligned.

Distribution is the spacing of elements so that they are equidistant. Distribution can be employed with features within elements, such as the point symbols in the legend in Fig. 13.5, or with a set of related elements, such as the elements along the right side of the page in Fig. 13.5. Most GIS systems have tools to distribute selected elements in the map layout so that they are spaced equally in either the horizontal or vertical direction.

Conclusion

This chapter reviewed the process of transforming geographic data into a map. This involves at least a basic understanding of the following: map projections and map scale; how to select, generalize, and portray that geographic information in graphic form; how to use symbols, color, and type on maps; how relief is shown on maps; and basic principles of map design and page layout. This chapter also described how cartography is a crucial part of GIS. From this chapter, you will hopefully have gained an understanding of how map use is integrally related to map compilation and design, and how effective map design cannot be achieved without keeping map use in mind.

References

1. MacEachren, A.M.: Visualization in modern cartography: Setting the agenda. In: MacEachren, A.M., Taylor, D.R.M. (eds.) Visualization in Modern Cartography, Pergamon Press, Oxford (1994), p. 1–12
2. Kimerling, A.J., Buckley, A.R., Muehrcke, P.C., Muehrcke, J.O.: Map Use: Reading, Analysis, Interpretation, 8th edn. Esri Press, Redlands (2016)
3. Snyder, J.P.: Map Projections: A Working Manual. United States Geological Survey, Washington (1987)
4. Battersby, S.E., Finn, M.P., Usery, E.L., Yamamoto, K.H.: Implications of web Mercator and its use in online mapping. Cartographica **49**(2), 85–101 (2014)
5. Dent, B.D., Torguson, J.S., Hodler, T.W.: Cartography: Thematic Map Design, 6th edn. McGraw-Hill, Boston (2009)
6. Jenks, G.F.: The data model concept in statistical mapping. Int. Yearb. Cartogr. **7**, 186–190 (1967)
7. Robinson, A.H., Morrison, J.L., Muehrcke, P.C., Kimerling, A.J., Guptill, S.C.: Elements of Cartography, 6th edn. Wiley, Hoboken (1995)
8. Bertin, J.: Semiology of Graphics. Esri Press, Redlands (2011)
9. Brewer, C.A.: Designing Better Maps: A Guide for GIS Users, 2nd edn. Esri Press, Redlands (2015)
10. Brewer, C.A., Harrower, M.: ColorBrewer 2.0 (2013). http://colorbrewer2.org/
11. Slocum, T.A., McMaster, R.B., Kessler, F.C., Howard, H.H.: Thematic Cartography and Geovizualisation, 3rd edn. Prentice Hall, Upper Saddle River (2009)
12. Imhof, E.: Positioning names on maps. Am. Cartogr. **2**(2), 128–144 (1975)
13. Imhof, E.: Cartographic Relief Presentation. Esri Press, Redlands (2007)
14. Schnabel, O.: Map Symbol Brewer (2007). https://www.oschnabel.carto.net/mapsymbolbrewer/

Further Reading

15. Allan, A.L.: Maths for Map Makers. Whittles, Dunbeath (2004)
16. Axis Maps: Cartography Guide (2020). https://www.axismaps.com/guide
17. Brewer, C.A.: Designed Maps: A Sourcebook for GIS Users. Esri Press, Redlands (2008)
18. Brown, A., Feringa, W.: Colour Basics for GIS Users. Longman, Harlow (2002)
19. Cauvin, C., Escobar, F., Serradj, A.: Thematic Cartography. Wiley, Hoboken (2010)
20. Dent, B.D., Torguson, J.S., Hodler, T.W.: Cartography: Thematic Map Design, 6th edn. McGraw-Hill, Boston (2009)
21. Kraak, M.-J., Ormeling, F.: Cartography: Visualization of Geospatial Data, 2nd edn. Pearson, Harlow (2003)
22. Krygier, J., Wood, D.: Making Maps: A Visual Guide to Map Design for GIS. Guilford, New York (2005)
23. MacEachren, A.M.: How Maps Work: Representation, Visualization, and Design. Guilford, New York (2004)
24. MacEachren, A.M.: Some Truth with Maps: A Primer on Symbolization and Design. Association of American Geographers, Washington, DC (1994)
25. Monmonier, M.: How to Lie with Maps, 2nd edn. Univ. of Chicago Press, Chicago (1996)
26. Snyder, J.P.: Flattening the Earth: Two Thousand Years of Map Projections. Univ. of Chicago Press, Chicago (1997)
27. Spence, M., Darkes, G.: Cartography: An Introduction. British Cartographic Society, London (2009)
28. Tufte, E.R.: The Visual Display of Quantitative Information. Graphics, Cheshire (1983)
29. Tufte, E.R.: Envisioning Information. Graphics, Cheshire (1990)
30. Tyner, J.A.: Principles of Map Design. Guilford, New York (2010)

Aileen Buckley Dr Aileen Buckley, a research cartographer at Esri since 2003, focuses on developing and sharing best practices for mapping. She publishes and presents widely on many aspects of mapping and GIS. She is a coauthor of Map Use: Reading, Analysis, Interpretation and a co-editor of the Atlas of Oregon. She holds a PhD in Geography from Oregon State University, and she was formerly on the faculty at the University of Oregon.

Paul Hardy Paul Hardy has an MA in Computer Science from Cambridge University, England. He spent 28 years at Laser-Scan in Cambridge, England, as Chief Programmer, Product Manager, and Principal Consultant, developing software for computer graphics and digital mapping. From 2003 to 2006, he was Product Manager for Cartography at Esri, implementing advanced cartographic capabilities into ArcGIS. From 2007 to 2016 he was Business Development Consultant for Esri Europe, advising national mapping and cadastral agencies. He has been since retired, apart from occasional consultancy.

Kenneth Field Dr Kenneth Field is a cartographer at Esri. He was formerly Principal Lecturer and GIS Course Director at Kingston University London, England. He holds a PhD in GIS from Leicester University and has presented and published widely. He is a Fellow of the British Cartographic Society and Royal Geographic Society, and he is a Chartered Geographer (GIS).

Geospatial Metadata

14

David Danko

Contents

D. Danko (✉)
Esri
Vienna, VA, USA
e-mail: ddanko@esri.com

Abstract

The concepts behind metadata have been around since the beginning of human communication, describing and recording what is known about objects/phenomena. This and the ability to communicate in detail – to interoperate – is what allowed humans to become the dominate species on Earth. Metadata is a primary interoperability enabler, providing information that is not apparent in the object itself: how old is it, when did it come into being, how was it made, where did it come from? Although the concept has been around a long time, the term *metadata* came about during the information technology (IT) age. Metadata is used to describe *data about data* or, more broadly, any information technology resource – data, services, knowledge stores, or other information types.

Geospatial metadata refers to IT resources that relate spatially to the Earth. It provides information which allows resources to be discovered by interested parties searching for resources, evaluated, accessed, and understood, enabling proper utilization. Metadata serves geospatial information IT resources throughout their lifecycle. It is both used and collected during the resource creation process. It is used to inform others of its availability and provide information to potential users as to the resource's pedigree and ensure it is fit for purpose. When employing resources and creating new knowledge, the metadata is used to back up the decisions made and help validate what was accomplished.

To truly enable interoperability, metadata must be widely understood. Users need to understand what metadata needs to be produced and what should be available to understand a resource. The metadata must be understandable by anyone using it, no matter their language or cultural

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_14

differences. This is what makes standardized metadata so powerful and important. To this end, many national geospatial metadata standards exist today as well as international metadata standards to promote global interoperability.

This chapter provides an introduction and background for metadata, defining the types of metadata and describing their uses. It also introduces international standards related to geospatial metadata and looks into the future of geospatial metadata and the advent of the Semantic Web.

Keywords

metadata · communication · geospatial · standard · interoperability

14.1 Background

We use metadata every day of our lives. We would never ingest medicine or packaged food without reading the label describing the product. Metadata, in some form, has always played a role in geographic maps, charts, imagery, text, and digital information. The earliest printed maps had legends that described the source of the map, the producer, the date, and usually the meaning of the symbols used. Today, of course, paper maps intended for serious users contain a wealth of metadata in their margins: information including title and other identification information, scale, reference system, accuracy, currentness, sources, producer, and extensive descriptions of the symbology. Many aerial photographs provided the metadata printed right along the edge of the photograph (Fig. 14.1). An aerial photograph containing the geographic location of the center point, the focal length, flying height, date, and time provides valuable geographic information; without it, it is just a picture.

Today, we download maps and directions from the Internet, relying on them, perhaps more than we should, without knowing much about them – typically, we do not know their origin or how up to date or correct they are. However, even here we use some intuitive metadata – if the map comes from

Fig. 14.1 A four-decade-old aerial photo with metadata in the top margin

Table 14.1 Metadata framework and stages of collaboration (adapted from [1])

Collaboration	Interoperability			
	Technical	Semantic	Legal	Human
Discover	Find each other	Determine meaning and usefulness	Determine business terms and requirements	Knowing where to look, understanding requirements
Trust	Establish physical connection trust	Establish information provenance trust	Establish business relationship trust	Understand agreements and intended use
Prepare	Prepare for physical data exchange	Prepare to use the information	Enter into a commitment with the other party	Take on the responsibility to receive and adhere to commitments
Transact	Exchange the physical data	Use the valid information	Establish transaction completeness and accountability	Adequate training to ensure successful results
Steward	Physically safeguard and protect the data	Safeguard the interpretations of the information used	Safeguard the organizations involved in the business relationship	Knowledge of organizational doctrine to properly manage the data

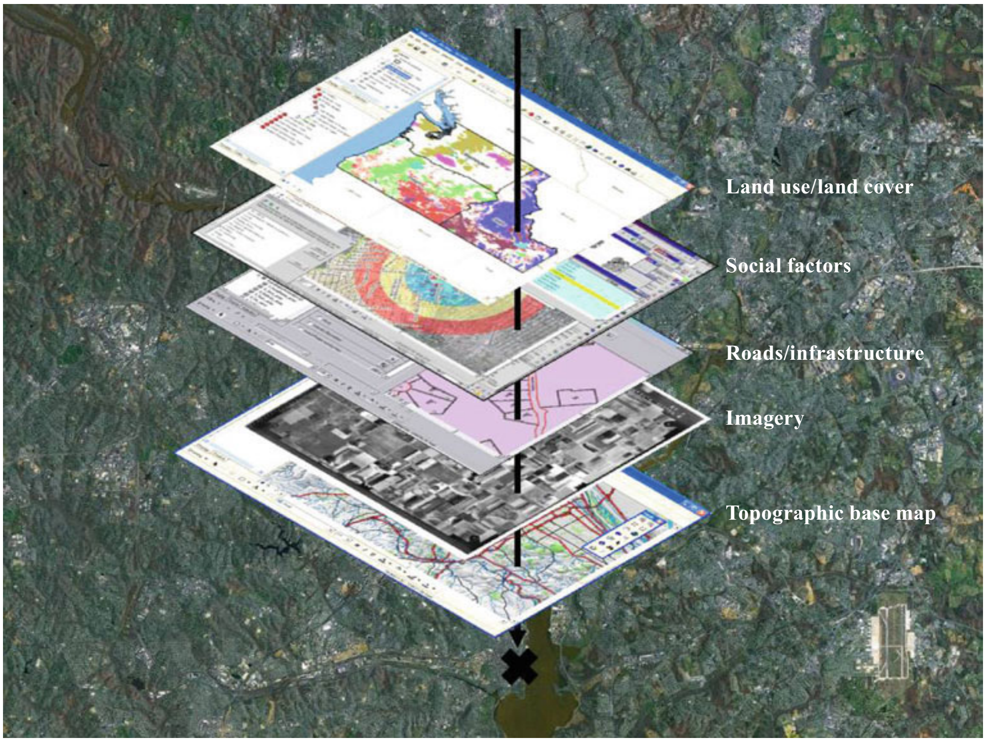


Fig. 14.2 Geography is a key integrating technology, tying diverse information types together through location

one of the navigation and mapping companies with which we have had good experiences, we tend to trust it, and we have come to learn what most of the symbols mean. However, if we are going to use maps and other geospatial data or services for meaningful, especially life or death, applications, we should demand adequate metadata.

Within the field of information technology (IT), metadata is data that describes the information so that it will be useful and have value, be understandable, and enable collaboration. Collaboration is interoperating purposefully toward a common end – the ability to share, exchange, use, and combine information (Table 14.1).

As we can see, collaboration is an activity that involves interoperability. Geographic Information Systems (GIS) have always required interoperability. Geography is the key that ties the information together in an information system (Fig. 14.2). When performing a geographic analysis, GIS operators use multiple sources from multiple organizations, often widely distributed. GIS merge diverse information types, merging feature data, imagery, computer-aided design (CAD) drawings, attribute tables, raster data, and more.

Interoperability requirements are becoming increasingly complex as GIS moves beyond isolated communities and into general information communities. Now, with the advent of geospatial web services and mapping, loosely coupled and operating over worldwide networks, interoperability is extremely important.

14.2 Interoperability

Interoperability is defined as [2]

> The ability of two or more systems or components to exchange information and to use the information that has been exchanged.

Interoperability goes beyond just the exchange of information; the information must be understood and usable following the exchange.

14.2.1 Categories of Interoperability

There are at least four major categories of interoperability [3].

1. Technical – this is typically the physical ability to exchange and use information. Examples are machine-to-machine connectivity, application program interfaces, the ability to read and understand formats, etc.
2. Semantic – this is the understanding of concepts and terms. When information is being exchanged, both parties in the exchange must have a precise understanding of the concept or term being exchanged. This category of interoperability is the most problematic, especially between disciplines, where often the same term may have very different meanings.
3. Legal/policy – this is the ability to legally exchange information, which is often controlled by intellectual property restrictions, export laws, ownership, and rights of use restrictions.
4. Human – this is the human capacity to enable an exchange and then to understand the information being exchanged. Interoperability is often thwarted due to a lack of cooperation and willingness to share. A lack of training or education often inhibits the ability to understand information in an exchange.

14.2.2 Interoperability Enablers

There are many enabling technologies, standards, and practices that enable these four categories of interoperability:

- Authorization to distribute or use information
- Copyright laws
- Business models and agreements
- Government and business policy frameworks
- Training
- Infrastructures
  - Networks
  - Hardware and software compatibility
- Information assurance
  - Certification of information
  - Quality and timeliness of information
- Support for multiple
  - Languages
  - Customs
  - Views
  - Data formats
  - Projections
  - Datums
- Metadata.

Metadata is important because it enables interoperability by contributing to an understanding of the information itself and of many of the above enablers. Metadata can provide copyright information, official authorization and terms of use, quality and timeliness of information, identify languages, customs, views, and describe formats, projections, and datums.

All of these enablers work much better if they are based on standards (Chap. 15). Standardized copyright laws, training, infrastructure, quality procedures and statements, formats, and metadata can be more widely and accurately understood.

14.2.3 Geospatial Interoperability

Interoperability, when it comes to sharing geospatial information, is more difficult because the world geography is so diverse and complex; geographic information can be provided in many different forms; and the fact that we all see things differently – because of our professions, our culture, and/or our view of the world. When we communicate our view of the world to someone else, this can be quite problematic, as there are many issues that need to be considered. When abstracting the real-world geography to produce geographic information, we develop a conceptual model to focus on the things we are interested in – our Universe of Discourse (Chaps. 1 and 15). When doing this, assumptions are made, some things are emphasized, and some things are left out, either by mistake or on purpose for clarification and to unclutter the model. To be fully interoperable, this view of the geography, the conceptual models, the abstractions, and the types of information that has been left out must be documented as metadata so that they can be understood by users of the information.

14.3 Applying Geospatial Metadata

Metadata is officially defined as *information about a resource* [4] – information that contributes to an understanding of a resource.

Table 14.2 Application environments

Environment	Application				
	Catalog	Processing	Production	Archival	Web service
Find	×		×	×	×
Evaluate	×		×	×	×
Access	×	×	×	×	×
Employ		×	×		

A resource can be in many forms: a service, an application, a web service, data, a database field, a physical object, an image, or just about any conceivable information technology item. Metadata allows us to find, understand, and use the right resource for the right reason, and use it correctly. This applies to geospatial metadata; most metadata element types that serve nongeospatial metadata are used to describe geospatial metadata as well. However, to describe geospatial resources, geospatial metadata must fulfill additional requirements. Examples are the geographic area to which the resource applies, abstraction methods, coordinate system, datum, positional quality, etc. These, and other aspects, will be fully described for each metadata element type below.

Metadata is used for a variety of applications; these can be classified into four primary functions: find, evaluate, access, and employ, which can be described in detail as follows.

1. *Find:* metadata enables users to locate geospatial resources and allows producers to *advertise* their data. Metadata helps organizations find data outside the organization and find partners to share in data collection and maintenance.
2. *Evaluate:* by having proper metadata elements describing a resource, users are able to determine its *fitness for an intended use.* Understanding the quality and accuracy, the spatial and temporal schema, the content, and the spatial reference system used allows users to determine whether a resource fulfills their needs. Metadata also provides the size, format, distribution media, price, and restrictions on use, which are also evaluation factors [5, 6].
3. *Access:* after locating a resource and determining if it meets users' needs, metadata is used to describe how to access it and, depending on the resource, transfer it to a specific site. Once it has been transferred, users need to know how to process and interpret the data and incorporate it into their holdings.
4. *Employ:* metadata is needed to support the processing and the application of geospatial resources. Metadata facilitates proper utilization of data, allowing users to merge and combine data with their own, apply it properly, and have a full understanding of its properties and limitations.

These four functions are performed in different environments, and the metadata must be provided in different ways to meet the requirements demanded by these environments (Table 14.2).

14.3.1 Catalog Environment

Typically, metadata used to find, evaluate, and access resources are provided for the catalog environment. Usually the metadata in this environment is provided separately from the resource itself. This metadata may be provided in both a human and/or machine-readable form. Hardcopy map and chart catalogs have served this purpose for centuries. One can use a catalog to search human-readable metadata text and graphics for maps and charts sorted by producer, type, scale, and geographic area covered. In addition to printed catalogs, users can now search for geospatial resources in geospatial clearinghouses and geoportals. A clearinghouse is a centralized physical or virtual structure for collecting metadata and other information about resources for the purpose of distributing and providing access to information about those resources. An example of a geographic clearinghouse is the Federal Geographic Data Committee (FGDC) clearinghouse [7]. A geoportal is very similar but is implemented as a web portal. Web portals are *web supersites*, gateways providing links and information for a wide variety of resources, typically addressing a specific theme. A geoportal is a web portal focusing on geographic information and resources. It typically provides an online metadata catalog to facilitate searching for geospatial resources, but it may also provide links to other information, calendars, community blogs, and reference information. A geoportal can serve as the organizing framework and key technology supporting spatial data infrastructures (SDI). Examples of SDI portals are the US Geological Survey (https://www.sciencebase.gov/catalog/) and the INSPIRE geoportal of the European Commission [8] (http://inspire-geoportal.ec.europa.eu/). Portals can range from comprehensive *home pages* for a specific theme or discipline, as in the two examples above, to more specific search portals. Examples of portals are Google and Amazon, and a geospatial example is Esri's ArcGIS (http://www.arcgis.com). Metadata catalogs are at the heart of all portal technology, whether comprehensive or simple search portals. All portals are based on a web services architecture where a service provider publishes metadata about a resource within a discovery service registry, and the user, or requestor agent looking for a service visits the discovery service to find a resource of interest. Once a desired service is found, the service requestor agent then interacts with the provider agent to access the service (Fig. 14.3; [9]).

A geospatial portal operates on the same principles. A geospatial data provider agent documents their resources as metadata (a web service description (WSD)) and publishes that metadata with a discovery service in a metadata

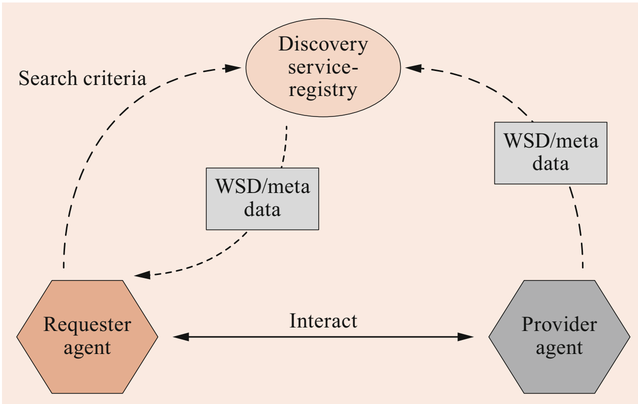


Fig. 14.3 W3C web services architecture – discovery (adapted from [9])

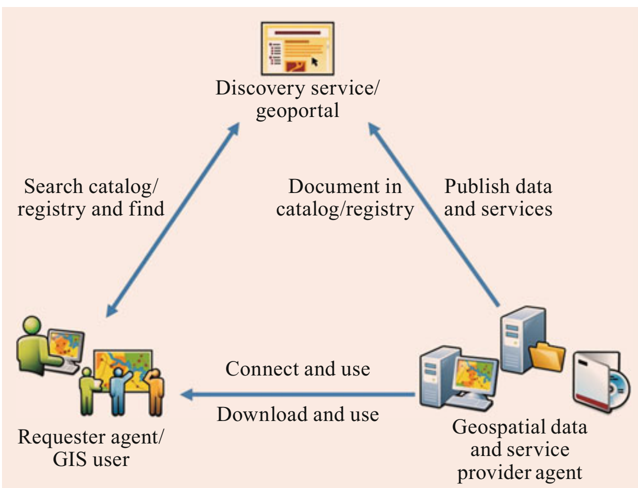


Fig. 14.4 Geoportal service architecture

registry/catalog on a geoportal. The resource provider can proactively publish metadata records on the geoportal or set up a harvesting agreement with the portal service broker, who will harvest [10] metadata records, which the provider agent exposes on the Internet at some prearranged place and schedule. Typically, this metadata is exchanged in the form of Extensible Markup Language (XML) documents. A GIS analyst or any user looking for geospatial datasets, services, or other resources visits the portal, searches the metadata catalog, and finds the data or service that meets his/her needs. If the geospatial resource he/she has found is a dataset, the link provided by the portal can be an order form where the user can order the data for download or physical shipment; if the resource is a web service, the link provided permits the user/requester agent to interact directly, providing instant access to the service (Fig. 14.4).

The metadata element types used in portals must support the user in identifying and *finding* potential resources; once found, the user needs to be provided with metadata that describes the resources adequately, so that they can *evaluate* the multiple resources found to be able to pick the resource that best fits their needs; once that is determined, they require metadata that describes how to *access* the resource. If it is a service, the access metadata could be a link directly to the service, or in the case of a dataset the metadata could contain ordering information.

As one can see, metadata is the fuel that enables portals, catalogs, clearinghouses, and any service architecture for that matter, to operate. If this metadata is standardized, then users, producers, and brokers across the globe will be able to better understand each others' metadata. Standardized metadata allow geospatial resource providers to know what metadata they should collect when documenting their resources and provides users with an understanding of what metadata may be available.

The more detailed the metadata available about a resource is, the more beneficial it is for the user. The ability to filter search results to eliminate unwanted *hits* is dependent on the quality and breadth of metadata available for searching and evaluation of resources.

Unstructured Versus Structured Geospatial Metadata

Google is the most successful search portal in the world today, and many argue that, since they do not require separate production of metadata for the resources they search for, structured/standardized metadata is irrelevant. There are three counters to this argument.

1. Google's business model is based on presenting interesting links for you to click on (they make their money from sponsored links), so they are intent on providing as many accurate links as possible but still want to tempt you with generally related sponsored links. Unstructured metadata serves this purpose very well. However, a scientific researcher or GIS analyst is interested in finding the best possible information to fill specific requirements and is responsible for evaluating the metadata for each resource to ensure its legitimacy for their research purposes.
2. Google does produce metadata for the resources they are cataloging; typically, the resources they are dealing with are Hypertext Markup Language (HTML) or other types of text pages from which metadata can be derived automatically. With geospatial data (consisting of long sequences of positions, values, and words but rarely containing structured text), on the other hand, the metadata is not so obvious or easily extracted. The metadata for geospatial imagery and vector datasets must be purposely collected near or at the time of the resource's creation so that the metadata is available for catalog searches, evaluation, and employing the resource. GIS researchers need to be able to search and understand many metadata variables that are

not inherently provided by the data, such as the time period of the content (not the date the file was created), positional accuracy, coordinate system information, etc.

3. Google does advocate for structured metadata. Google, Microsoft, Yahoo and Yandex have teamed to found Schema.org (https://schema.org) which has created a vocabulary and schemas for structured data on the Internet including webpages and email messages. Google's Earth Engine API (https://developers.google.com/earth-engine) is designed to take advantage of structured metadata if it is available.

14.3.2 Processing/Data Analysis Environment

As mentioned in Sect. 14.2, geospatial information is an abstraction of the real world. During the data capture process, a universe of discourse is created, focusing on the aspects of reality that are of interest to address a specific requirement. When GIS analysts are using and *employing* this data, they need to understand this abstraction. They need to know the purpose, point of view, and intended use for which the data were produced. Metadata needs to be available to provide this information. Typically, metadata supporting the employment of geospatial resources should be carried along with the resource so that it is always available to the user. A GIS analyst needs to fully understand the data model, e.g., how features are represented and whether vectors or raster/coverage schemas are used. If grid coverage schemas are used, what do the values assigned to the grid cells mean, and what is the resolution? If vector schemas are used, what is the level of topology, are edges connected, can network analysis be performed, at what resolution/level of detail are features provided; are cities represented by point features or area features; are city streets and buildings included? If the GIS analyst is working with multiple datasets, he/she needs a way to identify, keep track of, and understand the differences between them. He/she needs to know which has the best positional information or attribution information, and which is the most up to date. He/she needs the metadata to provide this information.

14.3.3 Production Environment

Metadata in the production environment helps in the management of the production process. Metadata here, in the form of extraction specifications, provides the individuals performing the data capture with an understanding of what to collect, to what level of detail, and to what accuracy. Here, we have metadata in the form of product specification driving the production, and we have metadata that records the results of the production processes and data sources. In this environment is ideally when most of the metadata for a resource should be produced.

14.3.4 Archival Environment

Most geospatial resources are expensive to produce and should be archived for future reuse. Historical geospatial resources are valuable for studying change over time. Geospatial resources, used in analysis, should be retained in case the analysis is ever called into question. Geospatial resources should be well documented with metadata to enable these possibilities. Metadata for archiving and preservation of geospatial resources should be carried with the resource and may be included in catalogs for search purposes.

14.3.5 Geoweb Services Environment

Metadata for geoweb services are usually provided in a machine-readable form, typically using XML in the form of service metadata or *capability* descriptions [11]. These descriptions are brief and define the name of the service, the type of service, the contained resource type names, if applicable, the geographic area covered by the service, a brief abstract describing the service, the intended computing platform, service operations metadata, service parameters and operations, and other web service-specific metadata. In this environment, the metadata supports discovery to *find* the service, and simple *evaluation*, and provides service metadata for *access*.

14.4 Economic Benefits of Geospatial Metadata

One of the benefits of metadata is that it allows producers to accurately identify the products they have produced, so that users and other data producers know they exist. Many times, if the existence of a resource is unknown, users who need data covering a specific area will collect data unnecessarily, remapping areas where data already exists. This is extremely costly and time consuming. It essentially doubles the cost – mapping an area twice.

14.4.1 Savings in Geospatial Resource Production and Management

Actually, a subset of metadata should be produced before the data is produced. When planning a production effort to collect geographic data covering a specific area, producers can publish metadata describing their production plans. Others interested in that same area can find out about it through the metadata and can team together, combining their efforts to share the cost. Many times, for example, one organization may want forest/land cover mapping, another wants transportation, and another the buildings; these three organizations can join together and map the area only once, by collecting the data to fill the requirements of all three organizations. This produces savings by having to do data collection (from aerial photography or ground surveys) only once; it will still cost slightly more, because the data collector will have to collect more information, but costs will still be saved by having only one collection. Collaboration networks can be established; when an organization is interested in collecting data in a specific area they can outline that area on a map and produce metadata identifying when and what type of information they plan to collect. That area with its metadata will be displayed on a map seen by others in the network; the others can outline their areas of interest as well. In this way, organizations can see, or be notified of, planned data acquisitions and can choose to share in the data collection.

Metadata is used to keep track of the status of data in production. Metadata is used to identify the source datasets to be used in production and to help in tying production data along *tie* lines. For instance, when performing quality control checks where two production datasets join, a road may be on one dataset but not on the other; by checking the metadata, you can learn that perhaps the source aerial photography for one dataset was taken at a later date than the source information for the other dataset. This allows producers to quickly understand the reason for the discrepancy, saving time and resources.

It is much more efficient and accurate to produce the metadata during data production. Trying to produce the metadata afterwards is difficult and may produce erroneous metadata. Using production software that automatically records metadata from production information, input parameters, intrinsic information, and data sources is much more efficient than capturing metadata by hand.

14.4.2 Geospatial Resource User Applications

When planning a study or project that involves analysis of geographic information, the first task is to identify and gather the geographic information to support the study. If metadata exists for all published geographic information, then the project planner only needs to discover and examine the metadata to identify the geographic information needed for the study and determine if it is fit for purpose. If the metadata did not exist, then the project planner would have no way of knowing what geographic data was available.

Efficient Resource Discovery and Evaluation

In an emergency situation, geographic source material must be located quickly. If the metadata for this source material is available in a catalog or online search engine, the metadata can be searched and the geographic information found quickly; for example, when planning an evacuation route for a flood disaster, the civil defense person needs to quickly find the terrain elevation data and the transportation routes to determine which routes will be safe to use for evacuation. If the metadata for these resources exists in a searchable form, they can be found quickly.

Metadata must exist to allow the potential user to determine whether the data is fit for their purpose. If the metadata carries information about the currency (the date the information is valid), the accuracy, the spatial resolution, and other metadata that allows the potential user to evaluate the data, they will be able to determine the correct datasets for their intended application. There may be multiple datasets covering an area of interest; the user must pick the dataset that suits his purpose as described above. Typically, datasets with higher accuracy and more dense detail are more expensive; if the user can use the metadata to determine which dataset has the minimum accuracy and the least detail that still suits his purpose, he may be able to use a dataset that is more cost effective. This may apply not only to the monetary cost of the dataset; a dataset that has too much detail may take more time and effort to generalize and rework. Conversely, if the user picks a cheaper, less dense dataset, it may cost more in the long run to enhance the dataset to suit his purposes. When proper metadata exists for datasets, the user can more easily determine the proper level of quality, currency, type of data (raster or vector), spatial resolution/density, and other evaluation factors to decide which dataset best fits his purposes.

Many times, geographic information is used to make legal decisions. The metadata for this geographic information should carry information about the temporal validity, the quality, the intended use of the data when produced, and any restrictions on its use; for instance, some datasets carry a warning in the metadata that it is not intended for navigation or commercial purposes. Having this metadata provides the backup documentation when legal decisions that depended on geographic information are being reviewed.

Efficient Resource Utilization

Once the potential user of a dataset locates the data and determines that it is fit for purpose, they must then know how to obtain the resource. The metadata should contain information so that the user will know: how to order the data or where to download it; the physical format of the data; whether it is in a format that they can use directly or whether it needs to be transformed; and if the data has a cost associated with it, the user needs to know how much and how/where to pay for the data. They also need to know if the data is copyrighted, or if there are any restrictions on accessing or using the data. The metadata should carry information about the security classification and legal restrictions on datasets. In this way, producers of catalogs and online search engines will know which geospatial resources *not* to include, or to only allow restricted access to these datasets. Metadata that documents the intellectual property rights or copyright restrictions of a geospatial resource and fully identifies the producer, owner, and the party responsible for the resource, helps ensure that the ownership is known and protected. The metadata benefits both the user and the producer by providing this information.

Once a user has used the metadata to find, evaluate, and access the geospatial resources they need for their purposes, the metadata should also provide them benefits when they are applying the data. The metadata is used during analysis to determine areas of the data that may be of lesser quality or temporally deficient. When using multiple geospatial resources during analysis they can use the metadata to identify and keep track of the various pieces of geographic information and determine which data should have priority during decisionmaking. When performing geographic analysis, the user needs to fully understand the data they are using. They need to know the data model, the application schema, the spatial schema (information about the geometry and topology), the temporal schema, the spatial resolution of the raster cells, or the vector vertices, the positional accuracy, the completeness, the attribute accuracy, the temporal accuracy, the logical consistency, the temporal validity, the use restrictions, the semantic meaning/feature coding/classification, the units of measurement, the format, the horizontal and vertical coordinate reference systems and datums, and information about the source, the lineage, and processes used to produce the information they are working with. All this information should be provided in the metadata about the data they are working with, so that they can properly apply the data, use it properly, and get the full benefit from it.

The automation of geographic information and sharing it within a nation is the major economic benefit of a National Spatial Data Infrastructure (NSDI):

1. It is ready and available to be used in a geographic information system.
2. It is easily distributed for others to use.
3. It can be utilized by multiple organizations simultaneously.
4. It can be applied in many different ways.
5. It is much more efficient and cost effective.

Metadata makes this possible and facilitates the use and application of digital geographic data. Metadata is the primary enabler of the spatial data infrastructure.

14.5 Geospatial Metadata Element Types

Table 14.3 presents the characteristics of the metadata element types that may be used to document resources [12]. Metadata types in italics are classes of metadata types followed by subtypes in plain text. This list is not exhaustive.

Table 14.3 Metadata element types

Metadata element type	Definition
Resource Identification information	Metadata that identifies the resource and distinguishes it from other resources and enables discovery
Title of the resource	This is usually the name by which the resource is known and can include the official title, alternate titles, identifiers, and ISBN information
Publication information	Information on the agent and when and where the resource was published or first made available
Responsible parties	The organization and/or position and/or individual responsible for the resource
Abstract	A brief description of the resource. A well-written descriptive abstract can be searched for keywords in context and can greatly enable search and discovery using text searches and provide a good description of the resource
Purpose	A description of the purpose for which the resource was developed (also considered a data quality metadata element type). By understanding the purposes for which a geospatial resource was created, users may better determine if it will fit their intended use
Usage	Information on how the resource has been used (if other than the intended purpose – also considered a data quality metadata element type)
Themes	Themes which describe the subject the dataset represents. Classifying geospatial resources by theme allows more accurate filtering of metadata for more precise discovery
Keywords	Search words. Plentiful and descriptive keywords enable a more intelligent search. Keywords from an identified thesaurus or ontology are even more powerful
Language	The language used in the resource, if used (some resources may not utilize text/human language)
Graphic sample	A simplified graphic enabling a quick evaluation on the resource's fitness for purpose. Typically this is a portable network graphics (PNG), graphics interchange format (GIF), or other reduced resolution overview of the resource. Often referred to as a "quicklook" or "thumbnail"
Status of and progress of the resource	Information on whether the resource is in production, finished, and/or available, or archived
Maintenance information	Maintenance and update plans and schedule. This is particularly important for geospatial resources used in maritime and aeronautical navigation where information is updated on a regular schedule and for data production managers

Table 14.3 (Continued)

Metadata element type	Definition
Spatial and temporal extent of resource	The extent of a resource which could be its extent of time, area, or vertical dimension
Spatial extent	The geographic area covered by the resource described by gazetteer/place name/terms or by geographic coordinates – bounding box, polygon or other geometric areas
Temporal extent	The time period covered by the resource (not the date of creation)
Content of resource	A description of the geospatial types used to create the resource
Coverage description	Description of the grid value types used (attributes) if the resource is an image or gridded data
Feature catalog description	A description of the feature types included if the resource is vector data
Quality of the resource	Metrics used to describe the quality of the resource; the measures of these metrics are reported as data quality measures [13]. Uncertainty categories. Lineage of the dataset
Positional accuracy	Accuracy of the geographic positions of feature/grid cells in the resource
Attribute accuracy	Accuracy of the descriptive attributes
Temporal accuracy	Accuracy of the temporal information
Completeness	Information regarding the degree of omission or excess of information in the resource
Logical consistency	Information regarding the resource's degree of adherence to its application schema; if the values are used in the values' domains; and stated topological characteristics
Lineage	Information regarding the sources, processes, and agents used to create the resource
Credibility	Factors such as reliability of the information sources, e.g., the source is correct 90% of the time [14–16]
Subjectivity	The extent to which human interpretation or judgment is involved in the information construction [14–16]
Interrelatedness	Source interdependences from other information [14–16]
Spatial and temporal types used by the resource	The identification and description of the data structures used to define the universe of discourse of the resource
Application schema	The identification and/or description of the application schema used (Chaps. 1, 4) by the data in the resource. Typically this is in a machine-readable, but may be in a human-readable, form understood by applications that are required to manipulate data that adheres to the identified application schema. Examples are a geography markup language (GML) application schema or Unified Modeling Language (UML) in graphical or XML metadata interchange language (XMI) form
Spatial reference system of the resource	A description of the reference system used to locate the resource relative to the Earth
Coordinate reference system description	The reference or complete description of the coordinate system used (Chaps. 6, 8)
Projection information	Information about the projection system, if used (Chaps. 8, 13)
Noncoordinate-based system description	Information about the spatial referencing by geographic identifiers that may be used by a resource. This is where a geospatial resource uses an indirect noncoordinate-based method for relating its data to the Earth using locations defined by a geospatial feature or features. Positions described by the relationship to feature(s) can be by containment – the data falls within a geospatial feature, e.g., a province, an ocean, a census tract; by local measurements from a feature where the position is defined relative to fixed points defined by a geographic feature, such as a measured distance from a road intersection; or more loosely as a position relative to geospatial features such as between two features. Typically, these feature-based geographic identifier systems are defined using gazetteers [13]
Availability and distribution of the resource	Information describing how the resource is distributed and/or how it is made available
Restrictions on access and usage of resource	Legal and security constraints for obtaining or using the resource (whether it is classified or whether there are copyright restrictions, etc.)
Cost	Fees required to obtain the resource
Format	The physical structure and information technology platform used to store or convey the resource
Distribution method	Where and how to obtain the resource
Service metadata	Information about a resource that is provided as a service
Service identification	Additional resource identification information (described above) that identifies the type of service and the capabilities that a service makes available to a user through a set of interfaces
Service operations	Describes the operations and parameters performed by the service
Data	Information identifying the data used or coupled with the service
Metadata about the metadata	Information that describes the metadata
Date	Information on when the metadata was produced, updated, or will be revised in the future
Standard used	The name and version of the metadata standard to which the metadata complies
Responsible party	The organization, position, or individual responsible for creating or maintaining the metadata

14.6 Applying Metadata to the Data

14.6.1 What Is a Dataset?

Typically, we think of metadata as applying to a unit of information commonly thought of as a dataset. A dataset is defined as an *identifiable collection of data* [4]; a digital aerial image, digital elevation data covering a national park, or a state road database, for example. However, a dataset could consist of a collection of one or more other datasets, which is sometimes defined as an *aggregate dataset*. Datasets can be aggregated for many reasons. Sometimes, a collection of datasets can be put together because they were all used together in a scientific or business study or other type of *initiative*. Datasets can be part of an aggregate or group of datasets because they are all aerial images from the same strip of photographs taken during the same flight with the same aerial camera. Sometimes, datasets can be part of a production group, all produced to the same product specification. The latter two types of aggregate datasets are examples of datasets belonging to a *series*. For instance, all the aerial photographs from the same flight are in a series because they are all from the same camera (sensor) and have the same equivalent focal length and other common specifics. Another example is *Serie Topográfica de Chile Escala* 1 : 50 000. This is a *production series* because all the datasets were produced to the same production specification. Figure 14.5 shows that an aggregated dataset can be a series or an initiative. A series can be a sensor series or a production series. An initiative can be a research study or a military planning mission.

Levels of Metadata

Metadata may be produced and be related not only to simple datasets but also to aggregate datasets, whether they are dataset series or datasets combined in a collection as part of some initiative. Metadata may also pertain to only specific parts of a dataset, such as specific feature types, property types, individual features, or individual feature attributes (Fig. 14.6). The specific levels of metadata are examined below.

Geospatial Dataset Metadata

Dataset-level metadata is the level to which most metadata records apply today. Most SDIs mandate metadata at the dataset level [4, 17, 18]. All existing and past metadata standards have mandated that each dataset shall have related metadata. Typically, the mandatory metadata records are focused on metadata for locating datasets, allowing the metadata to be exposed for discovery. As we have seen above, more comprehensive metadata allows dataset users to better understand and apply the dataset. Dataset metadata is best produced by the dataset producer, who has the best understanding of all aspects of the production process. If all features were extracted from the same source using the same process then the dataset metadata would be homogeneous. However, all data in a collection may not have the same properties, as some subsets of the dataset might be more current than others, be of different accuracies, or from different sources. In these cases, the metadata can contain related subsets that provide metadata about different data subsets related to different geographic regions or different feature or property types in the dataset; many times the dataset metadata is summarized, reporting the worst case, e.g., the earliest or average date when parts of the dataset have varying dates, the worst or average accuracy, etc.

Dataset Series Metadata

Metadata may be produced at the series level – a metadata file produced for each series – as well as for each dataset. A researcher looking for specific information can look at the metadata about a series without having to obtain the metadata for a specific dataset. They can look through the metadata for many series, find the series that contains the type of information in which they are interested, then look through the metadata for the many datasets, just in that series to find the one of interest. Metadata at the series level usually does not contain the full detailed metadata that dataset metadata contains. The series metadata should only contain the metadata that is common to all the datasets in the series, e.g., abstract, scale, topic category, distributor, etc. The series metadata may or may not contain the bounding box, unless the meta-

14

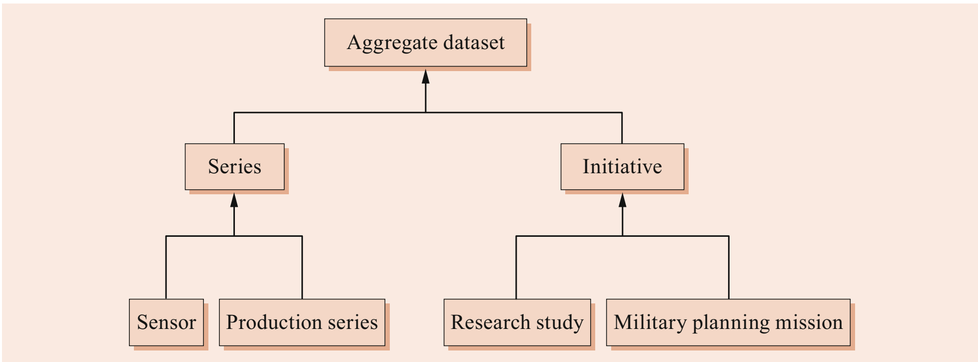


Fig. 14.5 Aggregate datasets

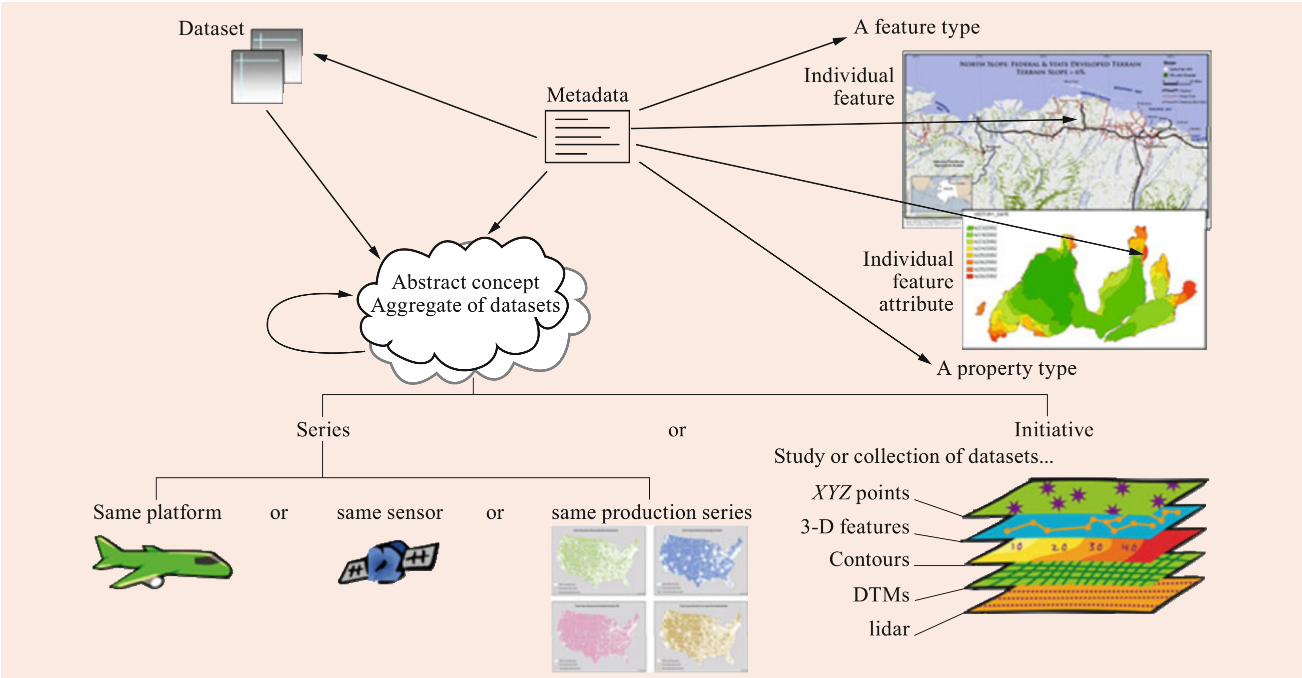


Fig. 14.6 Metadata may apply to different levels of data (DTM = digital terrain model; lidar = light detection and ranging)

data producer wants to put in the bounding box of the whole series. Series metadata can be used to make dataset metadata production more efficient; if the series metadata already exists – since it is common to all the datasets in the series – it provides much of the metadata for each dataset. All the dataset metadata producer has to do is to add the metadata unique to each dataset (bounding box, date of production, specific accuracies, etc.). The metadata may then exist and be provided in a hierarchy, so the series metadata is held at the series level, and the dataset metadata only carries the metadata unique to the dataset. The dataset metadata would require a key or link, pointing back to its series parent to provide a complete metadata record. The series metadata should not identify specific information about, or contain pointers to, the individual dataset metadata records.

Feature-Type-Level Metadata

As mentioned above, often different feature types in a dataset may be derived from different sources or using different processes; for example, vegetation feature types may be derived using an image classification processing step where the subjective polygon edges are determined through an automated process, whereas road feature-type positions may be precisely digitized from imagery and their attributes derived from a transportation database. In this case, it would be prudent to have feature-type-level metadata. Each feature type could be a separate dataset with separate metadata, or all feature types could be included in one dataset with a metadata record that contained subsets of metadata for the various feature types.

Feature-Level Metadata

In many instances, an individual feature, or all the features, within a dataset could have different metadata; for example, all the roads within a dataset were extracted from an image, but one road feature was added from data produced by a vehicle using a global navigation satellite system (GNSS) tracking and recording system. Typically, at the feature level, metadata is carried as an attribute of the features; for example, the road features in a dataset could have (data-related) attributes for number of lanes, surface type, speed limit, and coordinates, and a fifth (metadata-related) attribute describing positional accuracy.

Property-Type-Level Metadata

Metadata may be required at the property-type level; for example, a *tree* feature type may have a property (attribute) type *average height*, which is known to some level of accuracy. This accuracy could be provided as metadata about the *average height* property type.

Feature-Attribute-Level Metadata

As with individual features, the attribute of a specific feature could require separate metadata; for example, the attribute for the kilowatt power value carried by a power line feature may be classified as secret.

14.7 Geospatial Metadata Standards

As stated in Sect. 14.2, interoperability is greatly enhanced using *standardized* interoperability enablers – metadata being a prime enabler. A standard is defined as an agreement between a provider and a consumer, providing reference documents to be used in public contracts and international trade. Standards are definitions of characteristics, technical specifications, precise criteria, or rules that ensure that materials, products, processes, and services are fit for purpose, assuring quality, safety, reliability, and interoperability [19]. Many standards are established by nations for national benefit. International standards are developed for the same reason but for a wider audience – to promote interoperability between nations and around the globe. This section presents a collection of International Organization for Standardization (ISO) metadata standards. Many of these ISO standards can be used together to provide comprehensive metadata, or alone for a specific type of metadata, others provide a standardized encoding method (Fig. 14.7).

14.7.1 ISO 19115-1:2014 *Geographic information – Metadata – Part 1: Fundamentals*

This standard [4] defines the schema required for describing geographic information and services. The standard applies to all geographic resources – it is applicable to datasets in series, datasets, individual geographic features, and their attributes. It is also applicable to geospatial services. The standard defines the metadata required to serve the basic metadata functions such as resource discovery, determining a resource's fitness for use, and resource access, as well as optional metadata elements to support a more extensive description enabling a wide range of metadata applications.

Development

ISO 19115-1 was developed by ISO Technical Committee (TC) 211 – Geographic information/Geomatics and became an international standard (IS) in 2014. It is a revision and replaces ISO 19115, which was developed in 2003. It is designed to address the four uses for metadata described in Sect. 14.3 to facilitate finding, evaluating, accessing, and employing geospatial resources. Other national, regional, and special information community metadata standards existed prior to the development of the ISO standard; differences in terminology, structure, and purpose between these standards resulted in a lack of global interoperability. This incompatibility was one of the drivers that led to the development of the original ISO standard. However, the ISO standard was built on the experience gained in the development and use of these other standards, which were invaluable in its development. The initial ISO metadata standard was based on

- ANZLIC Working Group on Metadata: Core Metadata Elements, Australia and New Zealand Land Information Council, November 1995, Sydney
- Canadian Directory Information Describing Digital Georeferenced Data Sets, Canadian General Standards Board, July 1994, Ottawa
- Standard for Geographic Information – Metadata, European Committee for Standardisation (CEN), September 1996, Brussels
- US Federal Geographic Data Committee (FGDC) Content Standard for Geospatial Metadata, Federal Geographic Data Committee, 8 June 1994, Washington, DC.

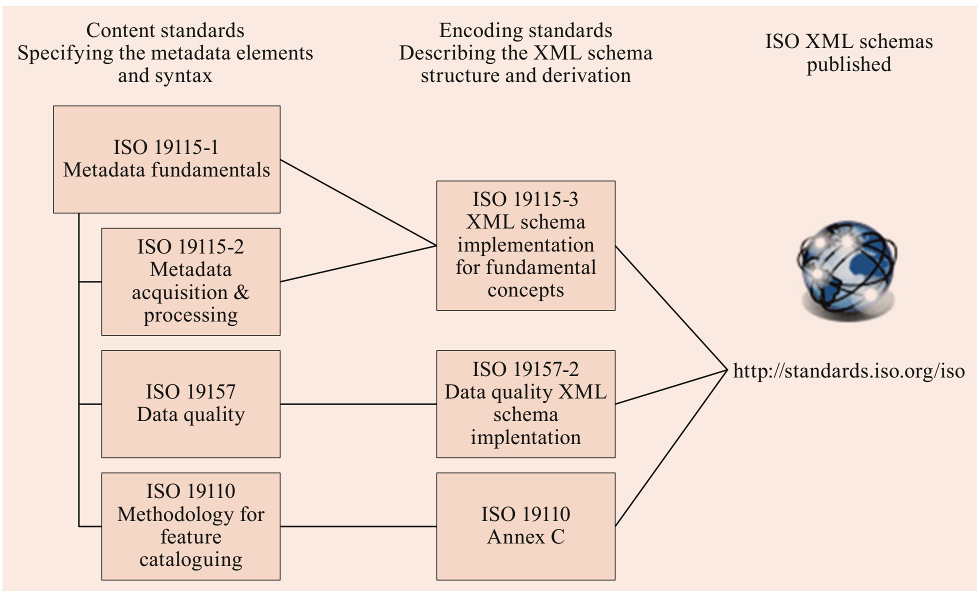


Fig. 14.7 ISO metadata standards relationships

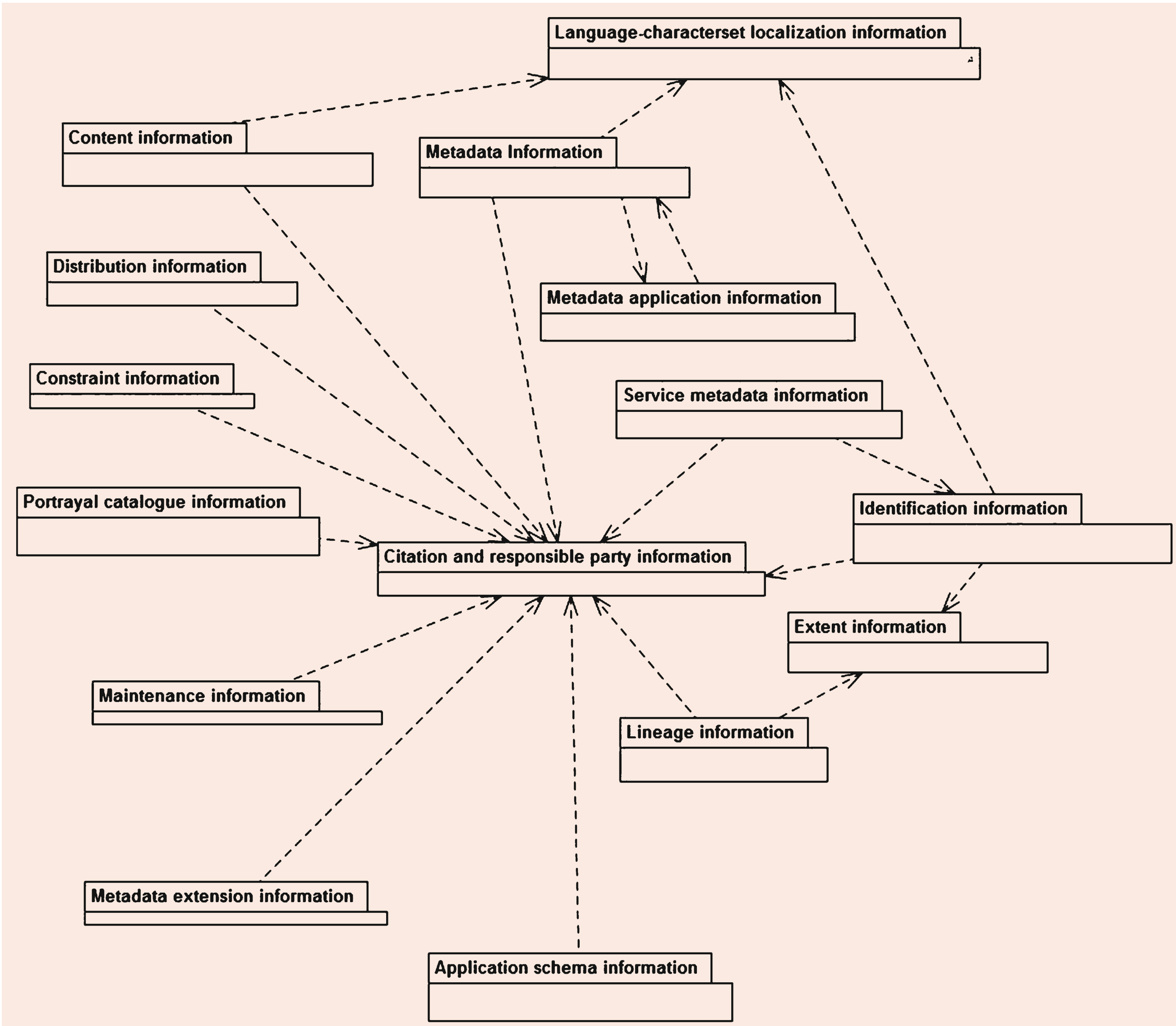


Fig. 14.8 ISO 19115-1 UML packages (adapted from [4])

Many transfer/exchange standards also carry metadata. These transfer standards were also examined and provided input to the ISO standard. Some of the transfer standards that provided metadata elements were

- *Digital Geographic Information Exchange Standard (DIGEST)*, Digital Geographic Information Working Group, January 1994
- *International Hydrographic Organization Special Publication 57*, International Hydrographic Bureau, October 1995, Monaco
- *Spatial Data Transfer Standard (SDTS)*, US Department of Commerce, August 1992, Gaithersburg, MD.

Experienced users of these standards also added input as to what worked, what did not work, and what was important. Project experts representing countries from around the world worked on both the original and the revised standards. These individuals represented the producer and user communities, so the metadata requirements and experiences of both communities were incorporated into the standards. The revised standard, 19115-1:2014, consists of the following parts under the title *Geographic Information – Metadata*:

- *Part 1: Fundamentals*
- *Part 2: Extensions for acquisition and processing*
- *Part 3: XML schema implementation of metadata fundamentals*

During the revision process metadata for services, derived from ISO 19119:2005 and ISO 19119:2005/Amd 1:2008 was

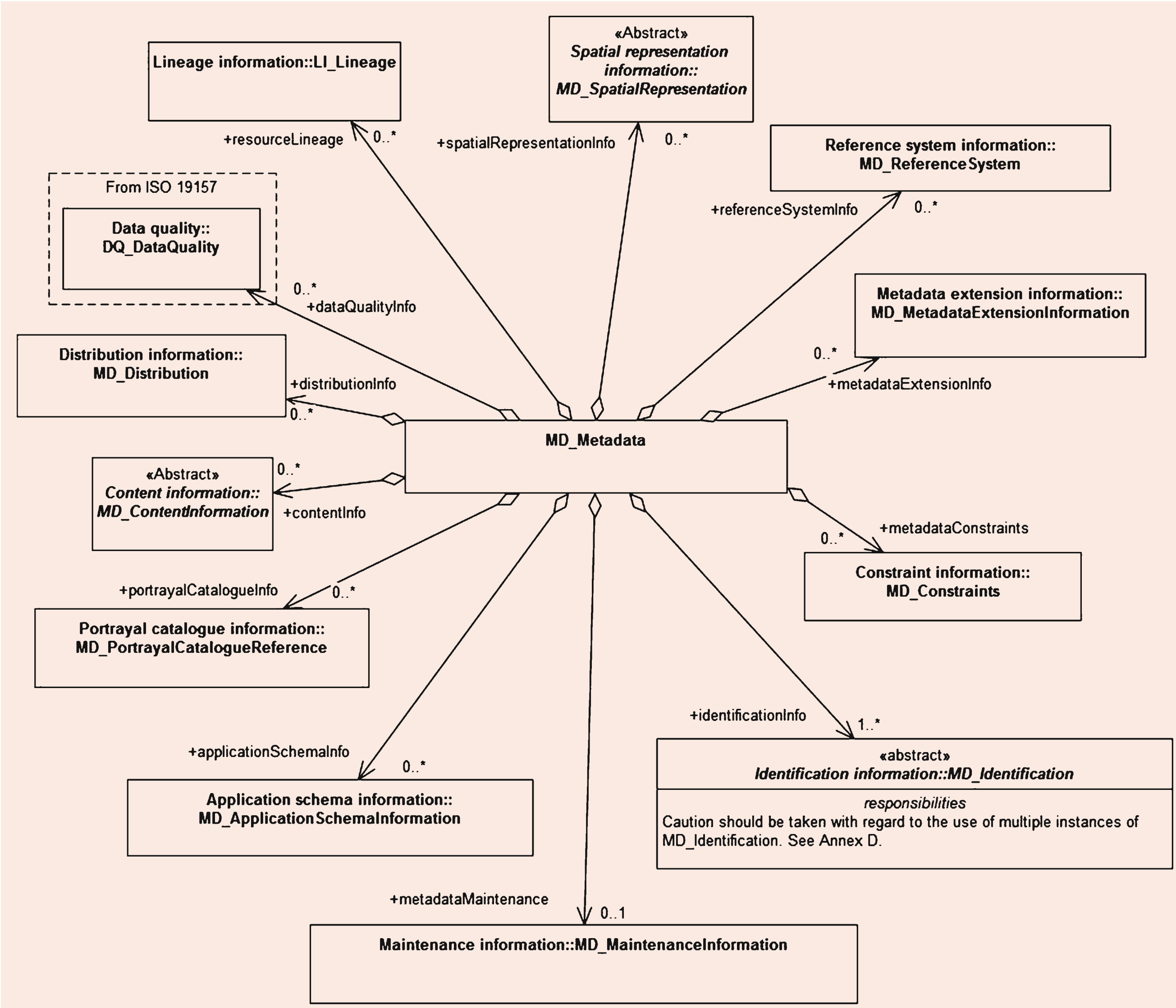


Fig. 14.9 ISO 19115-1 MD-metadata information UML model (adapted from [4])

added, and the data quality information was moved to ISO 19157:2013.

Description

ISO 19115-1 defines over 400 metadata elements to fulfill the requirements for finding, evaluating, accessing, and employing geospatial resources. These elements are defined in a data dictionary, based on *ISO 11179 Information Technology – Specification and standardization of data elements*, which provides a name, definition, obligation, maximum occurrence, data type, and domain for every element. It also provides a schema that defines their relationships using models based on the Object Management Group's (OMG) Unified Modeling Language (UML), which is also an ISO standard: ISO 19501.

The UML model is an integral part of the entire abstract model for geographic information developed by ISO TC 211 (Chap. 1). The metadata model is presented in 15 packages that combine collections of metadata classes and identify the relationships between these collections. Each package contains one or more classes of metadata, which may be specialized (subclassed) or generalized (superclassed). Each class of metadata contains attributes, which identify the discrete units of metadata. The UML packages and their relationships are shown in Fig. 14.8 [4].

Metadata specified in conformance with ISO 19115-1 consists of an MD Metadata class as the root class, which contains attributes (metadata) about the metadata and aggregates other metadata classes that provide metadata about geospatial resources (Fig. 14.9; [4]). This UML model spec-

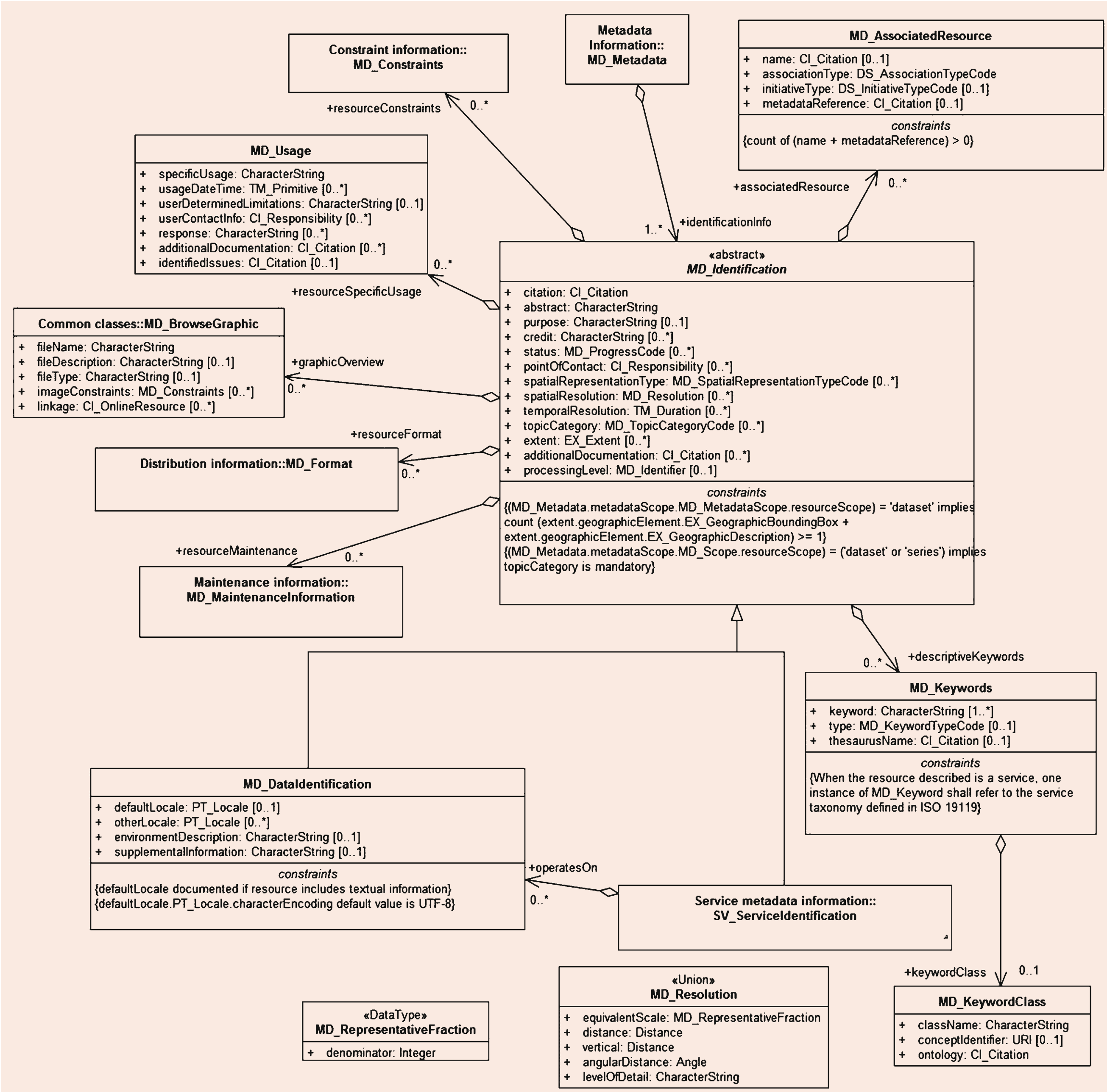


Fig. 14.10 MD_Identification UML model (adapted from [4])

ifies metadata about a resource must contain the following classes

- MD_Metadata – metadata about the metadata
- MD_Identification – a description that fully identifies the resource, including the spatial extent of the resource and service metadata if the resource is a service.

It also may contain

- MD_Constraints – identifying any legal or security, access or use restrictions associated with the resource
- MD_ReferenceSystem – identification of the coordinate or noncoordinate reference system used
- MD_SpatialRepresentation – a description of the geospatial information types used by the resource to digitally represent the Earth's features
- MD_Lineage – a description of the sources, processes, and provenance of a resource

- DQ_DataQuality – using ISO 19157 to provide a description of the quality of the resource
- MD_Distribution – a description of where and how to obtain the resource
- MD_ContentInformation – a description of the content of the geospatial information types in the resource
- MD_PortrayalCatalogueRefererence – a reference to the catalog of symbols/methods use to portray the resource
- MD_ApplicationSchemaInformation – a description of the machine-readable schema, which characterizes the structure and content of the resource
- MD_MaintenanceInformation – a description of the scope and frequency of updating and maintenance of the resource
- MD_MetadataExtensionInformation – a description of any nonstandard user-defined metadata included in the metadata for a resource

The standard further specifies properties, additional classes and subclasses for each of these higher level classes. Figure 14.10 provides an example of the level of detail specified in the MD_Identification class with the attributes, subclasses, and associations displayed.

These are similar to the metadata element types described in Sect. 14.5.

ISO 19115-1 Metadata Requirements

As shown above, ISO 19115-1 describes a schema and the definitions for metadata elements and identifies which elements are mandatory, which are conditional (and under what conditions they shall be used), and which are optional. It also defines which resources must be documented by metadata elements and the resources for which metadata may be required (Fig. 14.11; [4]). These resources are explained in Sect. 14.6.1 and Fig. 14.5.

Informative Classes

ISO 19115-1 contains additional information that provides guidance on the use and management of geospatial metadata.

- *Metadata extension methodology* provides advice on stages to follow when defining additional metadata elements.
- *Metadata implementation* provides guidelines for the management of metadata.
- *Implementation examples* are provided to help increase understanding of the standard.
- *Multilingual support for free-text metadata elements* provides a method for identifying the language of a free-text metadata element when the language for that element is different from the language specified for the dataset or when providing a metadata element in multiple languages.

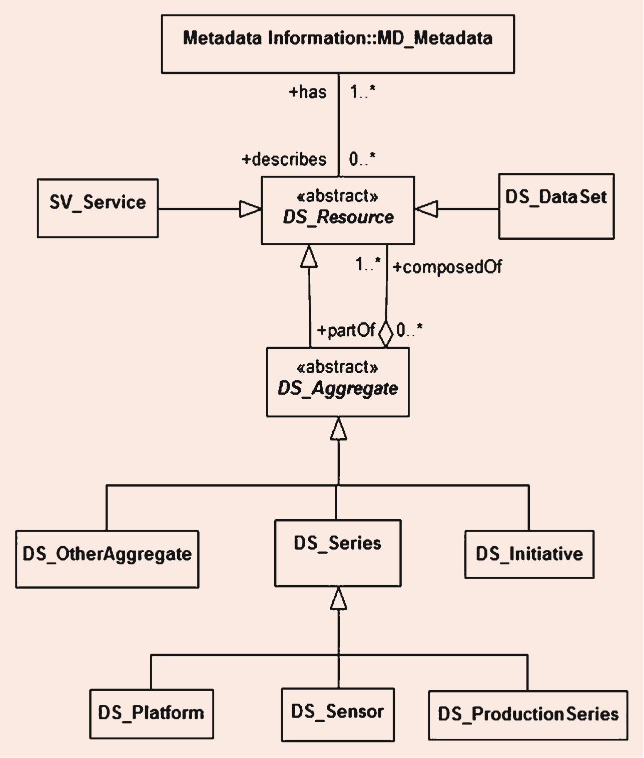


Fig. 14.11 Metadata applications (adapted from [4])

14.7.2 ISO 19115-2:2018 *Geographic Information – Metadata – Part 2: Extension for Acquisition and Processing*

ISO 19115-2 [20] is an extension of ISO 19115-1 [4] that defines additional metadata elements required to describe metadata about the acquisition and processing of geographic information, including imagery. This includes the properties of measuring equipment and the numerical methods and computational procedures used to derive geographic information from their output. It also includes metadata about the planning and development of requirements to produce the data, and extended descriptions of the processing steps used in refining the data. The standard extends the UML model and data dictionary of ISO 19115-1. ISO 19115-2:2018 *Geographic information – Metadata – Part 2: Extension for acquisition and processing* revises and replaces ISO 19115-2:2009 *Geographic information – Metadata – Part 2: Extension for imagery and gridded data*. During the revision of the standard it was determined that the acquisition and processing metadata in this standard applied to all geographic information resources, not just imagery and gridded data, hence the name change.

Description

As can be seen in Fig. 14.12, ISO 19115-2 extends the packages in ISO 19115-1, specifically, *spatial representation*

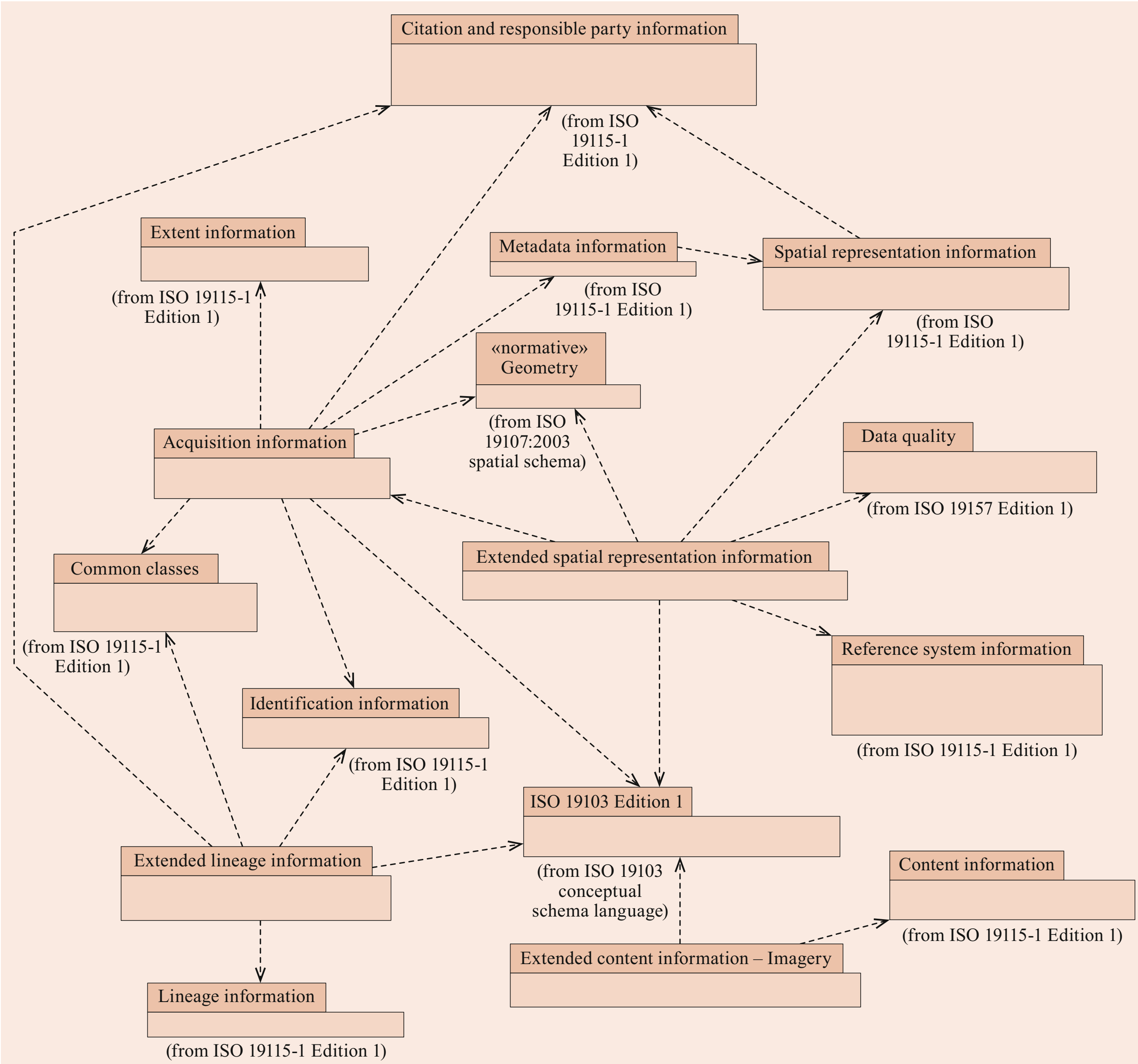


Fig. 14.12 ISO 19115-2 packages (adapted from [20])

information, *content information*, and *lineage information* and adds a new package *acquisition information* documenting the requirements, planning, measuring equipment, and other methods used in the acquisition of data.

The *spatial representation* information extension adds ground control points used for checking the quality of geographic information. These ground control points use data quality classes from ISO 19157:2013. The extended *lineage information* supports the description of additional source information, processing steps, parameters, and algorithms used in production of data. The *content information* extension adds additional classes to describe the sensors used in collecting the geographic information, primarily imagery. Additional imagery metadata, which supports calculating geographic locations from imagery using a physical sensor model, a true replacement model, or a correspondence model is defined in ISO/TS 19130 *Geographic information – Imagery sensor models for geopositioning* [21].

14.7.3 ISO/TS 19115-3:2016 *Geographic Information – Metadata – XML Schema Implementation for Fundamental Concepts*

ISO 19115-3 [22] provides a rule-based encoding for ISO 19115 Parts 1 and 2 in XML. It defines a collection of 18 XML schema namespaces roughly equivalent to the UML packages in ISO 19115-1 and 2. It also defines a set of XML schemas representing the aggregation of packages in ISO 19115-1 and 2 for various applications of metadata: metadata for data and services (a complete set of metadata for these two resource types); metadata application (for metadata resource collections with hierarchical metadata), metadata base for the basic mandatory elements, metadata for data transfer; metadata for data and services with geospatial common extensions; and metadata with extensions. The XML schemas were generated using the encoding rules from ISO 19118 (Chap. 4) and, where appropriate, they make use of GML and other previously defined XML encodings. ISO 19115-3 includes ISO/IEC 19757-3 Schematron rules that implement the validation constraints included ISO 19115-1 and 2 and also provides Extensible Stylesheet Language Transformations (XSLT) for transforming XML encodings of ISO 19115 using ISO 19139 XML schemas to ISO 19115-3 XML encodings and vice versa. ISO 19115-3 is a replacement for ISO/TS 19139:2007 *Geographic information – Metadata – XML schema implementation*. The ISO Metadata XML schemas are located on the ISO standards portal http://standards.iso.org/iso/19115/-3.

14.7.4 ISO 19157:2013 *Geographic Information – Data Quality*

ISO 19157 [13] defines the canons for describing the quality of geospatial data and specifies metadata elements for reporting the information. Geospatial data quality is defined as the characteristics of a geospatial resource that allow it to satisfy stated and implied needs. Data producers provide this data quality information stating how well a dataset represents a universe of discourse as defined in a product specification. Users can then evaluate the data quality information to determine a dataset's fitness for their application. ISO 19157:2013 replaces ISO 19113:2002 *Quality principles*, ISO 19114:2003 *Quality evaluation procedures* and ISO 19138:2006 *Data quality measures*.

Description

The standard establishes principles for describing data quality of geographic information by defining components for describing data quality; specifying the components and content structure of a register for data quality measures; describes procedures for evaluating the quality of geographic information; and establishes principles for reporting data quality.

Data is described using quantitative data quality elements. These elements are used to express how well data adheres to the criteria defined in a production specification. These values are typically results of a test or measurement, an evaluation of the difference between the data and the universe of discourse from which it was abstracted. The standard identifies six quantitative data quality elements, completeness, logical consistency, positional accuracy, temporal accuracy, thematic accuracy, and usability (Fig. 14.13), which are further classified into subelements.

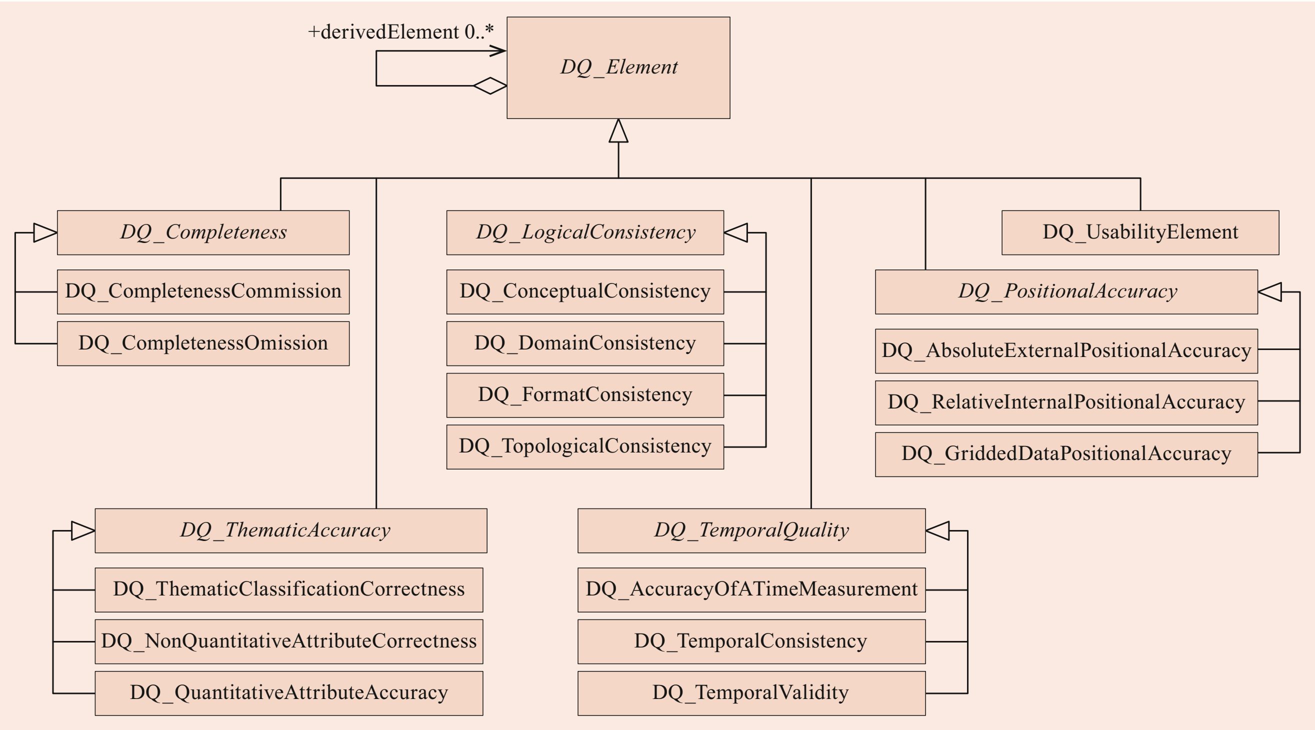


Fig. 14.13 Data quality elements

Completeness:

- Commission – excess data present in the dataset
- Omission – data absent from a dataset.

Logical consistency:

- Conceptual consistency – how well a dataset adheres to the rules of its conceptual schema.
- Domain consistency – how well values adhere to their value domains.
- Format consistency – the degree to which data is stored in accordance with the claimed physical structure of the dataset.

Positional accuracy:

- Absolute or external accuracy – closeness of reported coordinate values to values accepted as or being true.
- Relative or internal accuracy – closeness of the relative positions of features within a dataset.
- Gridded data position accuracy – closeness of a gridded data position values to values accepted as or being true.

Temporal accuracy:

- Accuracy of a time measurement – correctness of the temporal references of an item.
- Temporal consistency – correctness of ordered events or sequences.
- Temporal validity – validity of data with respect to time.

Thematic accuracy:

- Classification correctness – comparison of the characteristics assigned to features or their attributes to a universe of discourse.
- Nonquantitative attribute correctness – correctness of nonquantitative attributes.
- Quantitative attribute accuracy – accuracy of quantitative attributes.

Usability:

- Specific quality information about a dataset's suitability for a particular application.

These data quality metadata elements are reported as metadata in accordance with ISO 19115-1.

Data Quality Measures

The data quality measures provided in ISO 19157:2013 normalize the components and structures of data quality reporting, guiding producers in choosing the right measures when reporting dataset metadata and ensuring that users are presented with relevant and comparable measures when choosing the right dataset for their purposes.

It defines a set of commonly used data quality measures intended to be maintained in a register. Multiple measures are defined for each data quality subelement to support different types of data. The standard introduces the concept of two principal categories of data quality basic measures: *counting* and *uncertainty*. *Counting* related data quality measures are based on counting errors, while *uncertainty* is based on the concept of modeling the error of measurements with statistical methods. The standard lists 81 data quality measures. Examples of a counting data quality measure is shown in Table 14.4 and an uncertainty quality measure is shown in Table 14.5.

Using a registry, the data quality metadata will contain an identifier for the data quality measure linking to the register

Table 14.4 Standardized data quality measure – number of missing connections due to undershoots (adapted from [13])

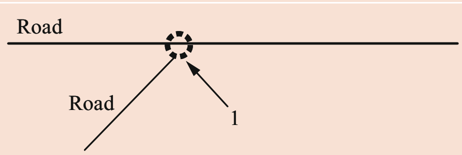
Line	Component	Description
1	Name	Number of missing connections due to undershoots
2	Alias	Undershoots
3	Element name	Topological consistency
4	Basic measure	Error count
5	Definition	Count of items in the dataset, within the parameter tolerance, that are mismatched due to undershoots
6	Description	–
7	Parameter	Search distance from the end of a dangling line
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	 Key 1. Search tolerance = 3 m
12	Identifier	23

Table 14.5 Standardized data quality measure – bias of positions (adapted from [13])

Line	Component	Description
1	Name	Bias of positions (1D, 2D, and 3D)
2	Alias	–
3	Element name	Absolute or external accuracy
4	Basic measure	Not applicable
5	Definition	Bias of the positions for a set of positions where the positional uncertainties are defined as the deviation between a measured position and what is considered as the corresponding true position
6	Description	For a number of points (N), the measured positions are given as $x_{\mathrm{m}i}$, $y_{\mathrm{m}i}$, and $z_{\mathrm{m}i}$ coordinates depending on the dimension in which the position of the point is measured. A corresponding set of coordinates, $x_{\mathrm{t}i}$, $y_{\mathrm{t}i}$, and $z_{\mathrm{t}i}$, are considered to represent the true positions. The deviation and biases are calculated as: Single deviations $e_{xi} = x_{\mathrm{m}i} - x_{\mathrm{t}i}$ $e_{yi} = y_{\mathrm{m}i} - y_{\mathrm{t}i}$ $e_{zi} = z_{\mathrm{m}i} - z_{\mathrm{t}i}$ Bias $a_x = \frac{\sum e_{xi}}{N_x}$ $a_y = \frac{\sum e_{yi}}{N_y}$ $a_z = \frac{\sum e_{zi}}{N_z}$ $a_{\mathrm{p}} = \sqrt{a_x^2 + a_y^2}$ $a_{3\mathrm{D}} = \sqrt{a_x^2 + a_y^2 + a_z^2}$ A criterion for the establishing of correspondence should also be stated (e.g., allowing for correspondence to the closest position, correspondence on vertices or along lines). The criterion/criteria for finding the corresponding points shall be reported with the data quality evaluation result
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	128

containing the full description of the measure. The components of a data quality measure are shown in Fig. 14.14.

Data Quality Evaluation

The standard defines processes for evaluating data quality (Fig. 14.15) throughout the phases of a product's lifecycle: specification, production, delivery, use, and update. It defines two evaluation methods: direct and indirect. The direct evaluation methods determine data quality by comparing the data to external or internal reference information. Using indirect evaluation methods data quality is determined using other metadata about the data, such as its lineage (Fig. 14.16).

14.7.5 ISO 19157-2:2016 *Geographic Information – Data Quality – Part 2: XML Schema Implementation*

ISO 19157-2:2016 *Geographic information – Data quality – Part 2: XML schema implementation* specifies a rule based XML encoding of ISO 19157:2014 *Data quality*. Like ISO 19115-1, ISO 19157 is a content standard specifying the data quality elements, data quality measures, processes, and methods for determining data quality, whereas ISO 19157-2 provides a method for encoding the information for machine to machine transfer. It provides the XML schemas in three namespaces; the XML schemas are located on the ISO standards portal: http://standards.iso.org/iso/19157/-2.

14.7.6 ISO 19110:2016 *Geographic Information – Methodology for Feature Cataloguing*

ISO 19110 [23] defines a methodology for cataloging feature types. It defines how a classification of feature types is organized into a feature catalog. A feature catalog is a catalog containing natural-language definitions and descriptions of the feature types, the features' attributes, operations, and associations provided as geographic data.

The XML schemas for ISO 19110 are located on the ISO standards portal: http://standards.iso.org/iso/19110.

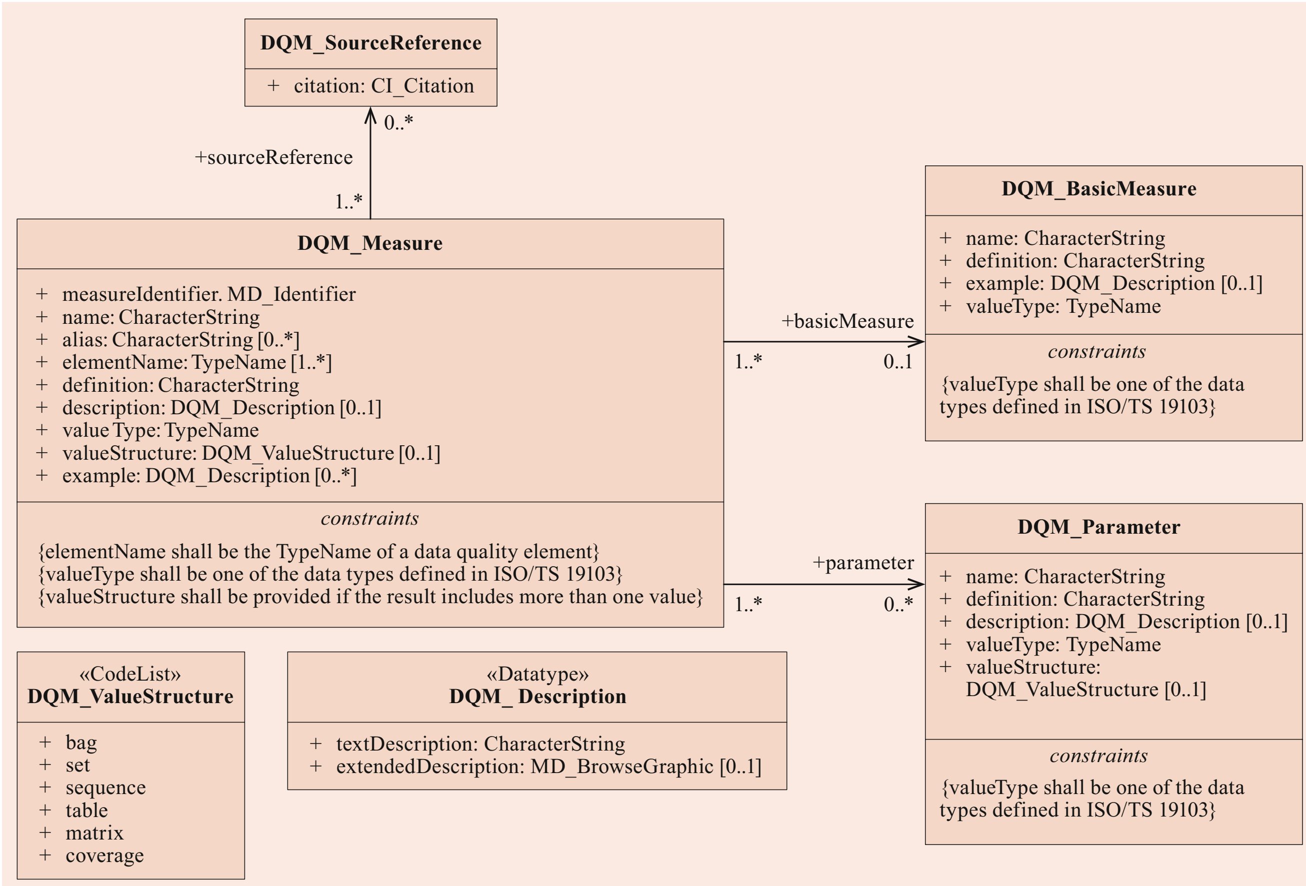


Fig. 14.14 Components of a data quality measure

As was mentioned in Sect. 14.2, individuals and organizations dealing with geographic information abstract the real world so it can be handled in modern information systems. When geographers form their universe of discourse, they identify the geographic features that they are interested in abstracting. Geographic features are real-world phenomena associated with a location relative to the Earth. Because individuals from different disciplines view or understand geographic features very differently, they must define how they see the world, describing their views by defining feature types using application schemas, data models, and feature and feature attribute catalogs; for example a logging company may view/define *trees* very differently than a landscape architect; a trucking company's view of a *road* versus a city planner's view, etc.

Feature types are classifications of geographic features grouped into classes with common characteristics. Application schemas and data models identify feature types by name, along with their attribution and relationships using graphical and machine-readable methods. As was mentioned above, feature catalogs contain definitions and descriptions of feature types along with descriptions of their attributes, operations, and feature associations using natural language, describing these feature types and their properties in more detail than can generally be provided in a modeling language.

A feature catalog consists of a collection of one-to-many precisely defined feature types, each with a unique (within the catalog) name and definition. Each feature type is characterized by zero or more property types, which can be a feature operation, a feature attribute, or an association role. A feature type can be a feature association whose role is described by an association role (Fig. 14.17). An example of a feature catalog can be found at https://inspire.ec.europa.eu/data-model/approved/r4618-ir/fc/.

Advantages of Feature Catalogs

Standardized feature catalogs provide metadata by providing a well-defined structure to describe the semantics of datasets. The purpose of a feature catalog is twofold. Firstly, it provides semantics of the data in datasets and as such can be used for data integration/conflation, reasoning, and analysis, an essential component of the direct use of the data; secondly, it provides the necessary information for discovery and assessment of fitness for purpose.

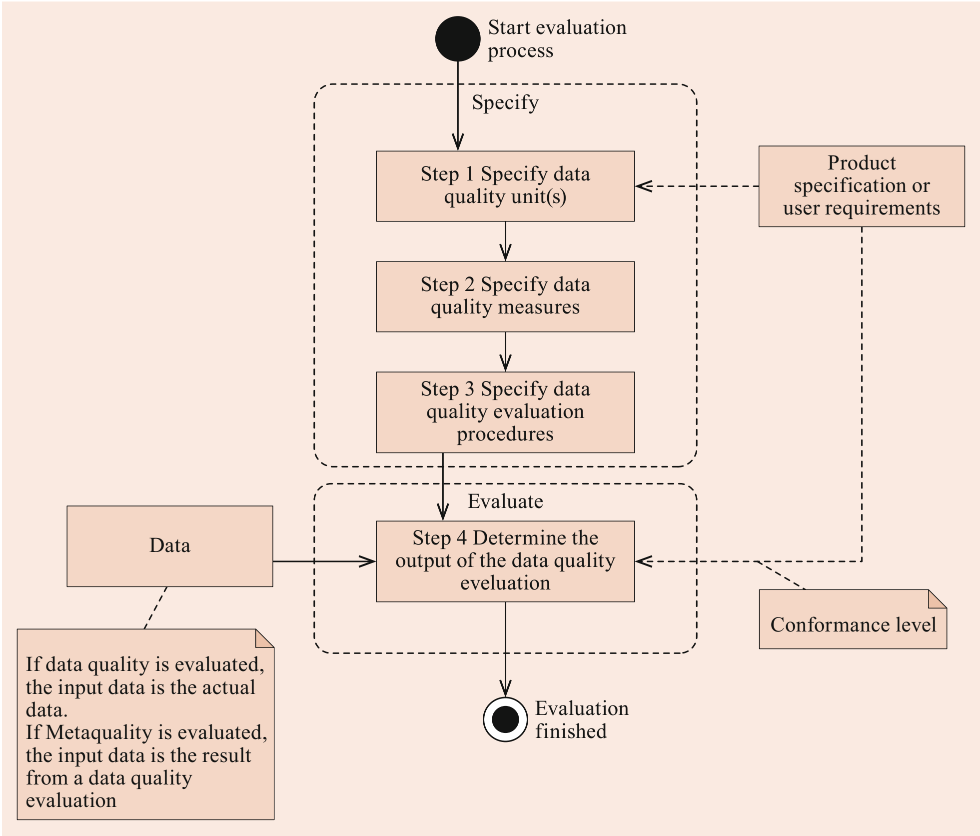


Fig. 14.15 Evaluating data quality

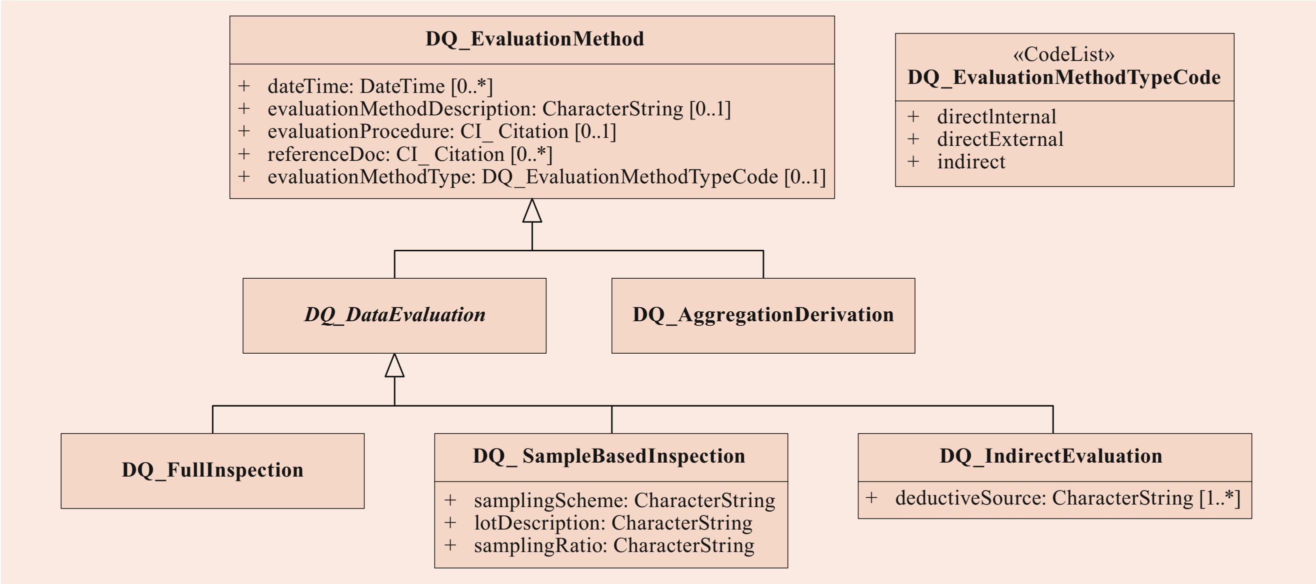


Fig. 14.16 Data quality evaluation methods

The advantages are that it:

- Fully documents a field of application's view of the world
- Provides a better understanding of the content and meaning of geographic data
- Ensures that geographic data can be fully and correctly understood and appropriately used
- Enhances communication between data producers and users
- Clarifies differences and helps harmonization between organizations that have overlapping domains.

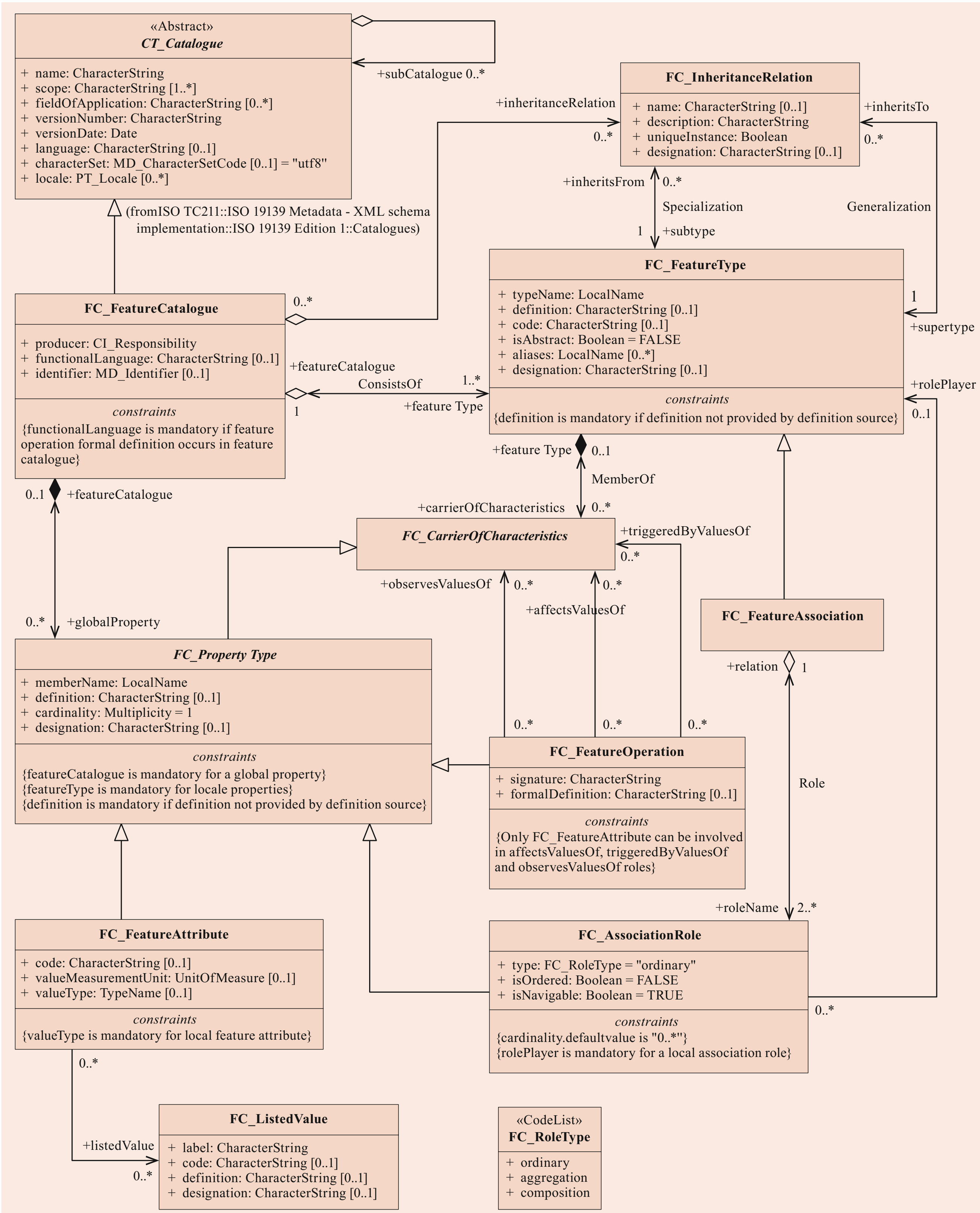


Fig. 14.17 Feature catalog conceptual model

Feature catalog metadata is used primarily in two ways.

1. A general catalog that applies to an entire collection of datasets that contains information. An example is the 1 : 100 000 scale digital line graph (DLG) hydrography and transportation series of datasets that cover the USA. A feature catalog would contain a detailed description of *all* the feature types, feature attributes, and associations that *may* be found in any of the 7687 datasets that make up the series. A feature catalog for this series provided to people producing the series tells them exactly what to digitize/include in the datasets (a product specification must also be used to provide them with minimum size, accuracy, and other collection criteria). The feature catalog allows users to fully understand the potential features types, feature attributes, and feature associations that may be found in the digital datasets in the series.
2. An instance feature catalog that is included with the metadata for each dataset, which lists *only* the feature types, feature attributes, feature operations, and feature associations found in that *specific* dataset. This aids search and discovery of datasets containing specific feature and attribute combinations. Examples include: find all datasets with feature type *roads* with *speed limit* attribute; purchase all USGS 1 : 24 000 DLG datasets containing *glaciers*.

14.7.7 ISO 15836 –1:2017 *Information and Documentation – The Dublin Core Metadata Element Set – Part 1: Core Elements*

ISO 15836-1 [24] defines a simple set of metadata elements for describing a wide range of networked resources that are designed for search and retrieval in catalog and web environments. Development of the standard began in 1995 in Dublin, OH, USA, hence the name, by digital library researchers, librarians and text markup experts. It became an ISO standard in 2003. Dublin Core was designed to document metadata about resources that does not necessarily have a geospatial component. Dublin Core metadata [24] is a simple set of metadata elements that can be easily produced and widely understood, and that promotes broad cross-domain interoperability (Table 14.6). It is designed to be used alongside – coexist – with other metadata standards that offer more domain-specific semantics (such as ISO 19115-1). It is often used with The Open Archives Initiative Protocol for Metadata Harvesting [10] for harvesting of metadata from data providers who expose their metadata to be used by service providers and who use the harvested metadata as a basis for building value-added services.

Table 14.6 Dublin Core metadata element set (adapted from [24])

Metadata element type	Definition
Title	A name given to the resource
Creator	An entity primarily responsible for making the content of the resource
Subject and keywords	A topic of the content of the resource
Description	An account of the content of the resource
Publisher	An entity responsible for making the resource available
Contributor	An entity responsible for making contributions to the content of the resource
Date	A date of an event in the lifecycle of the resource
Type	The nature or genre of the content of the resource
Format	The physical or digital manifestation of the resource
Identifier	An unambiguous reference to the resource within a given context
Source	A reference to a resource from which the present resource is derived
Language	A language of the intellectual content of the resource
Relation	A reference to a related resource
Coverage	The extent or scope of the content of the resource
Rights	Information about rights held in and over the resource

It provides common semantics for web metadata by defining 15 terms covering broad categories of metadata. There are no mandatory elements. As you may notice, every element has a comparable metadata element in ISO 19115-1, so common metadata for both standards can coexist in a metadata instance document, and one set of metadata can serve utilities that use either standard.

14.8 Geospatial Metadata Outlook

Geospatial metadata has always been important – and always will be in the future.

For centuries, metadata was carried in the margins of maps; the user has always had access to it. As we entered the digital era, metadata became separated from the data, carried as separate files, or held in production records. Many times, these metadata files, a minimum set serving only for the discovery of the information, only exist in catalogs. Professional mapping organizations capture comprehensive metadata; however, quite often, it is not made readily available. Users are making scientific decisions based solely on the reputation of the producer with no true knowledge of the quality and timeliness of the information. In today's world, anyone connected to the World Wide Web has access to, and the ability to upload and serve, information around the world. With the advent of GNSS and the ability of everyday users to collect positional information, many times as accurately as professionals (if they have the training and motivation to do so), ordinary citizens can participate in the production of geospatial information. Clearly, this volunteered geographic information (VGI) will require metadata so that other users can find it and trust it. Professional mapping organizations can ensure the production of metadata through standardized production procedures and employee incentives. With VGI, new uncomplicated, intuitive, natural methods for producing and managing geospatial metadata will need to be developed and assimilated into the production process.

With the advent of the Internet age, metadata has taken on new roles, new possibilities, new tools, and new uses. Early in the development of the Internet, it was realized that the future of metadata is the Internet, and the future of the Internet is metadata [25]. With the advent of Web 2.0, the web is becoming a bottom-up participatory sharing experience, mixing professional and nonprofessional-based information; methods have evolved through Wikipedia, flickr, ArcGIS, OpenStreetMap, Google Earth, etc., enabling the user to provide input. Methods for collecting user input metadata have evolved as well but will need to be expanded.

Metadata standards have been developed to precisely define geospatial metadata elements and their schemas in a language for geospatial professionals. Lay users/producers of VGI cannot be expected to use and understand these standards. User-centric metadata will have to be developed. To aid in the professional and scientific use of this information, it should be linked or cross-referenced to standardized metadata. This user-centric metadata should be in a form readily understood by users, provide a simple method for users to declare the information's fitness for use in their specific application, and enable them to add the lessons they have learned from using the information. It could also provide a web of data by providing many links [26, 27] to other related data and metadata. A set of OGC Geospatial User Feedback standards [28, 29] allow users of geospatial data products to provide feedback metadata such as ratings, comments, quality reports, usage reports, citations of related datasets or publications describing usage, additional provenance information, and significant events based on their experiences with using the data. They are designed to be used by metadata catalog servers and clients to manage and exchange user feedback information. This will allow users searching for geospatial information to meet their needs to be able to make informed decisions based on past users' input; like shoppers on www.tripadvisor.com and amazon.com.

14.8.1 Profiles

The ISO 19115-1 metadata standard has over 400 metadata elements, providing support for a wide range of metadata requirements. Only a few of the 400 metadata elements are mandatory. Typically, information communities will only use a subset of the standard. A profile is an adoption of a standard for a specific use [30]. Information communities will create a conformant subset of the standard and recast it in the vernacular of their community. Profiles of ISO 19115-1 provide a metadata standard tailored for use in a specific community such as a nation, a region, or an information community such as defense, hydrography, marine science, etc. Metadata elements are selected to meet the requirements of the community, code lists are extended or developed, and controlled vocabularies are established to use in place of specific free-text fields, and the document is written in a style readily understood by users in that field. As of this writing, several dozen profiles of ISO 19115 and Part 2 have been established or are under development, and many are being updated to be profiles of ISO 19115-1 and ISO 19115-2. An example of a regional metadata profile is the Inspire Metadata profile [18] developed for use by the European Union along with technical guidance for implementing the INSPIRE profile for datasets and services using ISO metadata standards [31].

Much smaller subsets of metadata for specific low-level applications, in addition to national profiles meant to support the wide-ranging needs of a nation, may also be established. For instance, a small subset/subprofile could be established in a particular application for data collection where providing the full conformant ISO, or information community profile, would be inconvenient. The individual collecting the data would only be required to collect a subset, for example, the name of the person collecting the data, the date and time of collection, and the position of the collecting instrument and other local conditions that could only be known at the time of collection. Of course, when the data is compiled and is provided as complete datasets, this metadata should be provided along with complete metadata fully compliant with the ISO standard or conformant profile. Conversely, an organization may produce much more metadata than required by the ISO metadata standards or a compliant profile. They may also store their metadata using more complex metadata element relationships than required by the standards to handle the management of data and metadata within their organization; however, when that metadata is presented to the outside world it should conform to the appropriate standards.

Two of the most important things a profile can do are to identify the fields/metadata elements that are important to the community served by the profile and to create controlled vocabularies for the words or phrases that are acceptable values for those metadata element fields. This will aid in the searching and understanding of the metadata within a community and support semantic cross referencing to improve searching and understanding of metadata across information communities.

14.8.2 The Semantic Web

The Semantic Web will be fully realized when methods and technologies are in place enabling machines to understand and process the meaning – the semantics – of the preponderance of the information on the World Wide Web. This will provide computers the capability to understand the meaning of the text on a webpage or the lines on a web map. Today's semantic methods primarily match words with some primitive intelligence using context, for example, discerning the meaning of the word *train* when it is used in coincidence with education as opposed to transportation-related words.

In the past, we have been using the web as a tool, chiefly HTML-based and not relying on database-driven websites, where metadata was created by web designers and software and not by website providers and users. Through Web 2.0, we have more control and interaction in our web experience, with users providing content, and more websites are database-driven and constructed and populated by user input and are highly customizable and utilize user-generated metadata such as user tags to enable search and understanding and to protect intellectual property. The focus of the network moved from one website with many users to many connected websites accessed by many users. However, much of the web is still an accumulation of unstructured information and, more importantly, unstructured metadata.

With Web 3.0, the Semantic Web, we are reaching the point where computers understand fully the semantics of the text on a webpage, for example, that a webpage about a *dog* is about a mechanical device holding/fastening a rope, not about a worthless person, or a canine. Or a web map entitled "Catoctin Mt." actually provides details on the spread of the marmorated stink bug, *Halyomorpha halys*, newly introduced into the USA from Asia, in orchards on Catoctin Mountain in Maryland between August 2009 and November 2011. Users can find and make proper use of this information in their research, applications, or protecting their apples.

The semantic web depends on ontologies (Chap. 17) and mappings between ontologies – a formal representation of knowledge by a set of concepts within a specific universe of discourse (domain) and the relationships between those concepts, a shared vocabulary, which can be used to model a domain – as well as having these ontologies available in a machine-processable form. This relies on a critical mass of metadata defining the information and providing metadata about the information. The idea of the semantic web is that when enough webpages carry machine-processable metadata, increasingly, developers will build tools to take advantage of it. For the semantic web to become widely available, webpage developers must be incentivized to provide machine-usable metadata. Web resources are identified by a Uniform Resource Identifier (URI) and some metadata; with just the URI the web has no way of knowing fully what the information is about, for example, when it expires, if it is suitable for children, or whether it supports specific research. Metadata is expensive; the last thing a webpage maintainer wants to do is fill out a lengthy form and continually update/manage the metadata, so much of it needs to be automated and easily added to webpages. The Semantic Web will be fully realized as metadata automation and management improves.

Today, most national and international geospatial data producers create geospatial metadata for the data and services they produce. Typically, they follow national or international geospatial metadata standards, which are, in fact, metadata ontologies for the geospatial domain. Increasingly, these organizations are using ISO 19115-3 Metadata – XML encoding to encode their metadata in a machine-readable form.

So, it seems that the geospatial community is ready for the Semantic Web. However, much geospatial metadata collected today focuses on *discovery metadata*, which focuses on a limited set of metadata to support users in finding geospatial datasets and services. Much of this metadata does not contain enough information to determine whether the geospatial information is suitable for the specific needs of the user searching for the information. Additionally, most metadata collected today does not contain information about the content/data model to provide an ontology of the geospatial information. The broader selection of comprehensive metadata element types defined in Sect. 14.5 is required to provide the semantics necessary to power the semantic web. Additionally, the production of metadata in the form of feature catalogs (feature and attribute definitions in XML, as GML application schemas, or other digital form) is necessary to provide a semantic understanding of the content of geospatial resources.

ISO/TC 211 has a suite of standards enabling the provision of these data and metadata ontologies, as well as encoding them in machine-processable forms; however, these are only understandable to those dealing with the geospatial domain. For the future integration of geospatial information into the wider Semantic Web, the geospatial domain will have to map its UML, GML, and XML encoded geospatial metadata and feature catalogs into Resource Description Framework (RDF), Web Ontology Language (OWL), Protocol and RDF Query Language (SPARQL), and other ontology technologies, and as well the broader web communities should be able to understand/utilize information based on geospatial community standards to bring the information into the Semantic Web.

The World Wide Web Consortium (W3C) provides and maintains most of the standards/recommendations enabling the Web. Data Catalog Vocabulary (DCAT) [32] is a W3C recommendation, which provides an RDF vocabulary used to describe datasets in data catalogs. It is designed to facilitate interoperability between data catalogs published on the web. This allows publishers to improve discoverability and allows application to consume metadata from multiple catalogs. DCAT makes use of Dublin Core for its data metadata vocabulary. The DCAT-AP is an application profile for data portals in Europe based on DCAT for describing public sector datasets in Europe [33].

To facilitate better integration between geographic data providers and the broader web community the European Commission has established the GeoDCAT-AP [34], which is an extension of DCAT-AP for describing geospatial datasets, dataset series, and services. It was developed to make geospatial data and services searchable on general portals (as opposed to geospatial specific portals). It does this by providing an RDF syntax binding for the union of metadata elements of ISO 19115:2003 (ISO 19115-1 in the future) and those in the DCAT-AP to provide a RDF based representation of geospatial metadata based on DCAT-AP vocabularies.

The W3C is also addressing geospatial interoperability with the wider web community with: *Spatial data on the web best practices* [35], which provides best practices related to the publication of spatial data on the web using Web technologies as they may be applied to location. These best practices are a significant change from the geospatial standards-based spatial data infrastructures approach to one based on general web standards. It describes 14 best practices; 8 through 14 are related to metadata:

- Best Practice 1: Use globally unique persistent HTTP URIs for spatial things
- Best Practice 2: Make your spatial data indexable by search engines
- Best Practice 3: Link resources together to create the web of data
- Best Practice 4: Use spatial data encodings that match your target audience
- Best Practice 5: Provide geometries on the web in a usable way
- Best Practice 6: Provide geometries at the right level of accuracy, precision, and size
- Best Practice 7: Choose coordinate reference systems to suit your user's applications
- Best Practice 8: State how coordinate values are encoded
- Best Practice 9: Describe relative positioning
- Best Practice 10: Use appropriate relation types to link spatial things
- Best Practice 11: Provide information on the changing nature of spatial things
- Best Practice 12: Expose spatial data through 'convenience APIs'
- Best Practice 13: Include spatial metadata in dataset metadata
- Best Practice 14: Describe the positional accuracy of spatial data.

Best Practice 13 recommends that descriptions of datasets that include spatial things should include explicit metadata about their extent, coverage, and representation. It makes note of using the GeoDCAT and metadata elements defined in ISO 19115 and promotes the use of GeoDCAT to specify spatial attributes not available in DCAT.

Another W3C set of metadata-related documents and standards, W3C PROV [36], addresses a web-centric model for providing information on the provenance of items and phenomena, not unlike the geospatial lineage described in Table 14.3 above. The PROV family of documents provides a model, definitions, and encodings for the interchange of information about entities, activities, and people involved in producing data or things. This will allow potential users make informed assessments about the quality, reliability, or trustworthiness of these data or things.

References

1. Sussman, R.: LIO-OPS Enterprise Metadata Strategy – As Mapped to Land Information Ontario (LIO). Land information Ontario, Hamilton (2003)
2. IEEE Computer Society: IEEE Standard Glossary of Softwere Engineering Terminology. IEEE, New York (1990)
3. Danko, D.: Geospatial metadata and interoperability. In: Shekhar, S., Xiong, H. (eds.) Encyclopedia of GIS. Springer, New York (2008)
4. ISO: ISO 19115-1:2014 Geographic information – Metadata – Part 1: Fundamentals. International Organization for Standardization, Geneva (2003)
5. Ahonen-Rainio, P.: Visualization of Geospatial Metadata for Selecting Geographic Datasets. Dissertation. Helsinki University of Technology, Espoo, p. 112 (2005)
6. Ahonen-Rainio, P., Kraak, M.-J.: Deciding on fitness for use: Evaluating the utility of sample maps as an element of geospatial metadata. Cartogr. Geogr. Inform. Sci. **32**(2), 101–112 (2005)
7. FGDC: Clearinghouse Concepts. Federal Geographic Data Committee, Reston (2009). http://www.fgdc.gov/dataandservices/clearinghouse_qanda

8. European Parliament: Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), Official J. Europ. Union L 108/1 (2007). http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF
9. Booth, D., Haas, H., McCabe, F., Champion, M., Ferris, C., Newcomer, E., Orchard, D. (eds.): Web Services Architecture. W3C Working Group Note 11. W3C, Cambridge (2004). http://www.w3.org/TR/ws-arch/
10. OAI: The Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 2.0 of 2002-06-14, Document Version 2008-12-07T20:42:00Z. (2002). http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm
11. OGC: OpenGIS Web Service Common Standard 2.0. Open Geospatial Consortium, Wayland (2010). http://www.opengeospatial.org/standards/is
12. Moellering, H., Aalders, H.J.G., Crane, A.: World International Cartographic Association, Spatial Metadata Standards. Elsevier, Amsterdam (2006)
13. ISO: ISO 19157:2013 Geographic information – Data quality. International Organization for Standardization, Geneva (2013)
14. MacEachren, A.M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., Hetzler, E.: Visualizing geospatial information uncertainty: What we know and what we need to know. Cartogr. Geogr. Inform. Sci. **32**(3), 139–160 (2005)
15. Roth, R.: A qualitative approach to understanding the role of geographic information uncertainty during decision making. Cartogr. Geogr. Inform. Sci. **36**(4), 315–330 (2009)
16. Bostrom, A., French, S., Gottlieb, S., Pang, A.: Visualizing uncertainty in natural hazards. Risk Gov. Soc. **14**, 261–294 (2008)
17. Clinton, W.: Executive Order 12906, April 11, 1994. Coordinating geographic data acquisition and access: The National Spatial Data Infrastructure. US Federal Register 59(71) (1994). http://www.archives.gov/federal-register/executive-orders/pdf/12906.pdf
18. European Commission: Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards to metadata. Official Journal of the European Union, L 326/12 – 326/30 (2008of). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32008R1205&from=EN
19. International Organization for Standardization (2021). http://www.iso.org
20. ISO: ISO 19115-2:2018 Geographic information – Metadata – Part 2: Extensions for acquisition and processing. International Organization for Standardization, Geneva (2018)
21. ISO: ISO/TS 19130:2010 – Geographic information – Imagery sensor models for geopositioning. International Organization for Standardization, Geneva (2010)
22. ISO: ISO/TS 19115-3:2016 Geographic information – Metadata – XML schema implementation for fundamental concepts. International Organization for Standardization, Geneva (2016)
23. ISO: ISO 19110:2016 Geographic information – Feature cataloguing methodology. International Organization for Standardization, Geneva (2016)
24. ISO: ISO 15836-1:2017 Information and documentation – Part 1:The Dublin Core metadata element set. International Organization for Standardization, Geneva (2017)
25. Iannella, R.: The future of Internet metadata. In: 5th Hong Kong Web Symp. (1999)
26. Berners-Lee, T.: Linked Data, Design Issues. W3C, Cambridge (2006). http://www.w3.org/DesignIssues/LinkedData.html
27. W3C: What is Linked Data. W3C, Cambridge (2010). http://www.w3.org/standards/semanticweb/data
28. OGC: OGC Geospatial User Feedback Standard: Conceptual Model. Open Geospatial Consortium, Wayland (2016)
29. OGC: OGC Geospatial User Feedback Standard: XML Encoding Extension. Open Geospatial Consortium, Wayland (2016)
30. ISO: ISO 19106:2004 Geographic information – Profiles Edition: 1. International Organization for Standardization, Geneva (2004)
31. INSPIRE Maintenance and Implementation Working Group: Technical Guidelines for Implementing Dataset and Service Metadata Based on ISO/TS 19139:2007. European Commission, Brussels (2017). https://inspire.ec.europa.eu/id/document/tg/metadata-iso19139
32. W3C: Data Catalog Vocabulary. W3C Recomendation 16 January. W3C, Cambridge (2014). http://www.w3.org/TR/vocab-dcat/
33. European Commission: DCAT for describing public sector datasets in Europe. EC, Brussels (2015). https://joinup.ec.europa.eu/release/dcat-ap-v11
34. European Commission: GeoDCAT-AP: A geospatial extension for the DCAT application profile for data portals in Europe, Version 1.0.1 (2016). https://joinup.ec.europa.eu/sites/default/files/distribution/2016-08/geodcat-ap_v1.0.1.pdf
35. W3C: Spatial Data on the Web Best Practices. W3C Working Group Note 28 September. W3C, Cambridge (2017). http://www.w3.org/TR/sdw-bp/#bp-metadata
36. W3C: An overview of the PROV family of documents (2013). https://www.w3.org/TR/prov-overview
37. ISO: ISO 19112 Geographic information – Spatial referencing by geographic identifiers. International Organization for Standardization, Geneva (2003)

David Danko David Danko is Senior Consultant with Esri, where he provides professional consulting services in GIS standards, interoperability, and geographic metadata. Previously, he worked for the US Defense Mapping Agency (DMA), where he was responsible for developing and managing agency policies and programs to ensure standardization and interoperability. David Danko has over 40 years experience in the fields of cartography, photogrammetry, remote sensing, and geographic information.

Standardization

15

Wolfgang Kresse, David Danko, and Kian Fadaie

Contents

W. Kresse (✉)
University of Applied Sciences Neubrandenburg
Neubrandenburg, Germany
e-mail: kresse@hs-nb.de

D. Danko
Esri
Vienna, VA, USA
e-mail: ddanko@esri.com

K. Fadaie
Retired from Department of Fisheries and Oceans Canada
Ottawa, ON, Canada
e-mail: kianfadaie@gmail.com

Abstract

This chapter contains a detailed presentation of the geomatics standards of the ISO/TC 211 *Geographic information/Geomatics* and the Open Geospatial Consortium (OGC) as of 2021. Section 15.1 is dedicated to interoperability, which is a driving force for standardization along with many other interoperability enablers, such as quality assurance, confidentiality, integrity, availability, authenticity, nonrepudiation, usability, verification, validation, and metadata, which provide the trust and understanding of geospatial resources necessary ensure interoperability. The standards are categorized as de jure and de facto standards. The standardization organizations ISO, IEC, and ITU are presented in Sect. 15.2, and the standardization procedure, from working draft, via committee draft to international standard is outlined.

Section 15.3 describes the development of ISO/TC 211 from its predecessor, CEN/TC 287 *Geographic information*, to the present program of work. The most basic standards, the reference models, are explained, followed by a discussion about standards implementation strategies. A roadmap to the complete 19100 standards family is shown in the beginning of Sect. 15.3.3.

Section 15.4 illustrates the manifold of links from ISO/TC 211 to other ISO committees and international organizations as internal and external liaison members, respectively. Each of the liaison members is briefly portrayed. Section 15.5 is dedicated to the OGC, which is

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_15

an industry standardization body for implementation level standards. The structure and the standards development process are explained in detail. The standards are grouped into interoperability specifications (IS) and abstract specifications (AS); the latter served as a template for some early ISO 19100 standards. OGC standards are continuously being extended, and many of them have been forwarded to become ISO standards. Each AS and IS is explained in detail. A glossary of ISO-terms with about 1000 entries is closely linked to this chapter.

Keywords

standard · ISO/TC 211 · Open Geospatial Consortium · Interoperability · ISO · OGC · CEN

15.1 Interoperability

Interoperability is achieved when varied persons, organizations, or systems can work together. This is important for geographic information because of the requirement to share information and processes in a broad field involving diverse information types about the globe with its diverse and complex features and processes engaging a wide spectrum of applications, people, and organizations. Data from around the world are defined and collected by humans of different cultures using a variety of sensors and methods; these data are used by a wide variety of processors operated by many types of organizations. There are three principles that must be met to realize interoperability:

1. The ability to find what you need when you need it
2. Once located, the ability to access and obtain what is needed
3. And after obtaining it, to be able to understand it and put it to good use.

Standards play a major role in providing interoperability but are by no means the only factor.

When performing geographic analysis we need to produce the information needed or find information produced by others, access and bring the information into a system, merge it with other information, and perform high-quality, coherent analyses that can be easily and accurately understood by others.

15.1.1 Infrastructure

Infrastructure is critical to interoperability. Infrastructures are critical in the use of geographic information. A spatial data infrastructure (SDI) consists of the technology, the coordination between organizations and agencies, the policies, a supportive environment for the sharing and utilization of spatial data, and international standards for interoperability of resources [1]. A clearinghouse that collects and distributes spatial resources is an example of an activity supported by an SDI. Clearinghouses are repositories that collect, store, and disseminate resources or metadata about resources and can be physical or virtual [2]. A multiperspective description of an SDI is described in [3]. SDIs exist at the national, regional, and global levels. Examples are the US National SDI [4], the one of Latin America [5], the infrastructure for spatial information in the European Community (INSPIRE) [6], and the Global SDI [7].

15.1.2 Training, Knowledge, and Human Resources

Interoperability is only possible if the information is provided, accessed, and used by knowledgeable people and organizations. As one of the tenets of interoperability is putting accessed information to good use, users of information in an interoperable exchange must have the necessary training and knowledge to understand and use it properly. Users must know, for example, which type of resources can be used for specific analysis. The military, for example, has found that data at a specific level of detail, accuracy, and currency, and an understanding of the reference system, are critical in specific mission planning and combat operations, and has established extensive training on official doctrine for the use of geospatial information. Geographic information system (GIS) software vendors provide extensive training in the use of their products to ensure proper geographic analysis and application of their products. Many universities provide undergraduate degrees, multidisciplinary graduate education, and advanced research programs in the field of geospatial information science [8]. GIS may be applied in many fields. Efficient and proper application of GIS in these fields requires education and knowledge in information technology, databases, and in some cases Web applications, geodesy and coordinate reference systems, remote sensing, statistics, and spatial theory, as well as the subject where the GIS is applied: environmental protection, military, facility management, business analysis, etc.

15.1.3 Information Assurance

Interoperability does not work if the information cannot be trusted or is unfit for purpose. Information assurance (IA) is a formal concept/practice developed by the US defense and intelligence agencies to manage risks in the use, processing, storage, and exchange of information/data [9, 10]. These concepts are just as important in the transmission, storage, and utilization of, and decision-making using, geographic information. Information assurance includes the following items:

- *Quality*, which is defined in ISO 9000 as (the) "degree to which a set of inherent characteristics fulfills requirements." Information or any exchanged resources must be of sufficient quality to fulfill the requirement for which they were obtained for interoperability to be worthwhile. Producers need to ensure that their products are of suitable quality for their intended purposes and publish that quality information as metadata. Information about the quality as metadata allows users to select the geographic resources that are fit for their purposes, furthering interoperability.
- *Confidentiality*, which means preserving the restrictions on information access and use, protecting privacy and proprietary information to ensure that information is not disclosed to unauthorized entities. Interoperability may not be realized unless the parties in the exchange know that their information will be protected through copyright, secure handling procedures, and other means.
- *Integrity*, which means ensuring that data/information has not been changed, altered, or improperly modified in an exchange to ensure information nonrepudiation and authenticity.
- *Availability*, which means timely and reliable access to geographic resources to ensure the usability of the resource and confidence in the outcome of analysis performed using it.
- *Authenticity*, which means the property of being genuine, verifiable, and trusted to enable the geographic resource to be used with confidence.
- *Nonrepudiation*, which means ensuring that the sender of information is provided with proof of delivery, and that the recipient is provided with proof of the sender's identity, so neither can deny having processed the information. Uses of geographic resources that are backed by the provider ensure confidence in the use of the resource.
- *Usability*, which is defined by ISO as *the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use*. With respect to geographic data quality, ISO 19157 [11] defines it as (the) *degree of adherence to a specific set of data quality requirements*.
- *Verification and validation*. Verification is ensuring that a resource is designed correctly to address specific requirements, while validation ensures compliance with that design. So, verification is done infrequently: at the time of resource design/development or when determining its fitness for use, whereas validation is done continually to ensure that the resource consistently meets the verified requirements. For example, verification of a system ensures that the geospatial data it uses and the processing performed can determine regional flood potential. The data input is then validated, so that it meets the verified requirements.
- *Authoritative data*, which are officially recognized data that can be certified and are provided by a legal authority or that adhere to a specification and doctrine for their use. Land cadastre is not legally binding unless derived from authoritative data [11, 12]. In many cases, the user does not need to know the detailed accuracy and timeliness information, as long as they use data authorized for a specific application.

15.1.4 Metadata

Metadata, defined as information about a resource, are a major interoperability enabler because they describe the information (to be) exchanged. Metadata are a primary interoperability enabler because they address all three of the principles of interoperability. Metadata in the form of a brief description of a resource that can be indexed for searching enable the resource to be discovered, and if the quality of the information of the resource is included in the metadata, potential users can determine whether the resource will fit their needs. Once a resource has been selected, metadata that provide information about cost along with access and use restrictions and provide information about the structure and format of the resource allow users to obtain the resource. Metadata that allow users to fully understand resources enable them to properly put it to good use (Chap. 14).

15.1.5 Standards—De Jure, De Facto, Industry

A standard is a documented agreement between a producer and a consumer, i.e., a reference document to be used in contracts and international trade, which specifies definitions of characteristics, technical design or content, precise criteria, rules, or guidelines. Standards ensure that material products, processes, and services are fit for purpose. Standards are typically focused on a specific community/user group and region. Examples are international standards (all regions), regional standards (for trading blocks such as the European Committee for Standardization), national standards (American National Standards Institute (ANSI), Deutsches Institut für Normung (DIN), etc.), community standards (Society of Automobile Engineers (SAE)), company standards developed for use within a single organization, and government standards focused on specific national and local government needs and applications. An example of international standards focused on a specific maritime community are those of the International Hydrographic Organization (IHO), which are focused on the safety and efficiency of maritime navigation such as the *S-4 Chart Specification of the IHO* and *Regulations for International Charts* and *S-52 Specifications for Chart Content and Display Aspects of ECDIS* (Chap. 23). Interaction amongst international organizations plays a major role in the greater understanding of roles and functions and promotes technical cooperation, which results in greater interoperability. For example, the technical cooperation between the International Maritime Organization (IMO), which provides for regulatory carriage of nautical charts through its Safety of Lives at Sea (SOLAS) Convention, greatly promotes the enhancement of hydrographic surveys and the dessimination of the data. Similarly to a commercial product, to be successful, a standard must be well designed to fulfill its purpose, maintained and kept up to date, and well publicized.

As explained below, standards may be de facto or de jure. De facto standards, defined as *in fact, whether with legal right or not*, are product practices or standard techniques which have become dominant in a market. De facto standards can be unofficial or compulsory (de jure) standards that are dominant in a market when one or more standards exist for the same use. De facto standards exist because they were the first, best marketed, or the simplest path to interoperability. The QWERTY keyboard is an example of a de facto standard. Being designed to be inefficient to slow the operator down so that keys would not jam (less frustrating) along with other factors to make the typewriter easy to use (single keys versus dial and key, inked ribbon, and roller bar for loading the paper) and used by the two most popular early manufacturers, it became dominant in the market and is still the de facto standard today, even though better designs have been proposed [13].

De jure standards, defined as *by law, legally accepted*, are standards mandated by an organization or nation and usually endorsed by a standards development organization, which are the subject of the sections on standards below.

15.2 Basics of Standards

Although everybody recognizes that standards have become essential in every corner of our life, they are usually considered a dry-as-dust topic that seems to block inspiration and flexibility. Standards are assumed to be a weak compromise between existing and proven solutions developed by dull administrative people.

15.2.1 Characteristics of Standards

In reality, standardization is a real challenge, often comparable to the most sophisticated development projects in the industry. Good standards are simple and unique; for example, Roman letters are in use after more than 2000 years, and they are still able to adapt to almost any language of the world. In contrast, Roman numerals were not that successful as a standard, because the theory behind them was not mature enough. As everybody knows, a numbering system without a zero has limited applications. But once developed, standards are by nature inflexible and there is reluctance to move from one to another.

All of us have gained some experience with standards during our lifetime. An example of a standard is A3 and A4 paper sizes used in most parts of the world, or the letter and legal paper size standards used in America. This kind of standard is helpful, and today, nobody would start arguing about the standard paper sizes.

Another group of examples are file formats. The Microsoft Word *.doc format is well known. Images are often stored or transferred in the tagged image file format (TIFF). We use them regularly, or rather we let our computer use them. These formats have been developed in conjunction with particular computer programs. The *.doc format belongs to the Microsoft Windows word processing system. TIFF was created by Aldus and adopted by Adobe.

These examples illustrate the vast and heterogeneous nature of standardization. Standards can be technical or management oriented, detailed or abstract, the result of an international consensus-building process or of a single company's development.

To move closer to the world of standardization, the subject can be viewed from various perspectives.

Linguistic Perspective and Types of Standards

In Medieval Europe, a standard was a flag or sculptured object with the distinctive ensign of a king raised on a pole to indicate the rallying point of an army. It was also used later for the authorized example of a unit of measure or weight [14]:

- In modern English the word "standard" has several different meanings, even within the standardization business: An official standardization organization such as ISO publishes formal standards that have a certain level of relevance in the application domain. Often, this type is called a *de jure* standard. In the French and German languages, the corresponding terms are "la norme" and "die Norm".
- Companies develop industry standards for the purpose of operating their products. These standards are not officially branded unless they put them through the official standardization process. If the standard is well accepted by the user community, it becomes a so-called *de facto* standard.
- In sectors such as the information technology (IT) business and geomatics, consortia of companies agree on common specifications to ensure interoperability. These

specifications are also *de facto* standards. However, in recognition of the companies' intention to develop common technical rules, they are called standards from the beginning.

Economic Perspective

The drive for industry to invest in standardization efforts is to benefit from the tremendous financial savings that ensue. The reasons in detail are:

- Standardization avoids the costs of adapting interfaces to a range of applications.
- Participation in the standardization process puts companies ahead of others not participating in the process.
- Standards enable a company to utilize a range of suppliers rather than becoming reliant on a limited number of sources.
- Standards support the legislation process. About 20% of the German standards (DIN) are referenced by laws and by-laws and take the burden of solving detailed technical questions away from the legislative body. In this sense, standardization simplifies the legislation process, as parliament does not have to deal with the subject in full detail. Abstract laws need only refer to the DIN standards that cover the details.

User's Perspective

If a technology is mature, then users expect standardized solutions that are simple, fast, and effective; for example, in 1996 a production company looking for a data exchange format that suited all practical needs for airborne imagery type data initiated the International Society for Photogrammetry and Remote Sensing (ISPRS) working group (ISPRS WG) dedicated to standardization. This user-driven case is a typical starting point for standardization.

After the initiative had started, discussions in the ISPRS WG and other committees revealed that a standard had to be designed in a much broader sense than originally intended.

A well-known format for referencing imagery to the Earth is GeoTIFF. It is a good solution for small-scale or georectified imagery but has some strong limitations. GeoTIFF is based on a well-defined set of TIFF tags. The original idea was simply to extend the set of TIFF tags to meet the requirements of airborne photogrammetry. However, it turned out that a simple definition of additional TIFF tags would only foster a temporary solution, as new sensor types continuously show up on the market. As programming time is expensive, one would not decide for interim solution. A desirable extension of the subject towards generic transformations and general sensor geometries is not possible without having developed an extended theoretical foundation and strategies to adapt the implementation to the latest solutions with minimum effort.

In fact, the modern solutions are model-driven, based on the ISO 19100 standards. The implementation has been using the extensible markup language (XML).

System Manufacturer's Perspective

The best solution for a system manufacturer is a closed universe based on the manufacturer's system and without any interfaces to the outside world. Today, this is no longer realistic. Government agencies and other customers have forced manufacturers to open their systems and support standardized interfaces. In a continuously expanding market such as geomatics, it seems that sharing resources with industrial partners can favor the economic growth of a company.

If standardization is pushed towards an implemented solution, the companies not only have to pay for drafting the standard's documentation but also for programming the implementation. This can be expensive. To guarantee a return on investment, companies in the IT business (including geomatics) have opted for joint solutions. The best-known example is the Open Geospatial Consortium (OGC).

Development Group's Perspective

A development group, usually called the standardization committee, must draft a document that contains a solution for the standardization task. The document should also address future challenges, the details of which are often unknown when the development takes place. In this sense, standardization, particularly in the IT domain, is very closely related to the design of computer systems. A lot of money is invested according to the content of a specification, and the better the specification, the longer the investment will be enjoyed.

Choosing the right moment to launch a standard development is probably the most critical decision in the process. If it starts too early, newer industrial developments will overtake the standardization documents. If it starts too late, there might be little room left for standardization among the existing solutions that have established dominant positions.

As standardization generally takes place *after* industry has completed a number of developments, the standardization committee is always confronted with existing solutions, mostly with limited compatibility. Most software companies are not willing to invest in software adaptation, because modification of existing software towards compliance with a new standard is extremely expensive. This has accelerated the creation of industrial consortia in the IT domain, where common industry standards are often defined before the investment in software development starts. The development of a sufficiently generic standard in combination with the ability to define profiles for specific fields of application is a widely used strategy to integrate existing software into new standards. This strategy leads to compatibility at the abstract level but not necessarily at the implementation level. Often, extended compatibility can only be reached with future versions of the application software or with the advent of newer technologies. An example is the extensible markup language as today's *de facto* standard for encoding documents and database contents.

National Perspective

The principles of standardization are: do it once, do it right, and do it internationally. In the IT domain, any national approach is at best a preparation for or a benchmarking of an international solution. A typical example of this process is the history of the European approach to the standardization of geomatics through CEN/TC 287 *Geographic information* (CEN, Comité Européen de Normalisation). This committee became dormant after the ISO/TC 211 started its work, and its members' expertise and knowledge were integrated into the new worldwide committee. However, CEN/TC 287 was reestablished in 2003 to confirm the ISO/TC 211 standards as European legislation.

However, international standards do not mean that national or regional distinctions are overridden. An important subject is the linguistic adaptability of standards. As the English language dominates world business, in particular in the IT sector, it is often difficult to maintain the coherence of a standard-compliant system and a national user interface. Admittedly, this is often a matter of costs (Chaps. 1 and 4).

Differences in traditions are more difficult to solve. The first North American GIS were designed for medium-scale applications. From the beginning, they permitted integration of remote-sensing technology. European systems followed their cadastral tradition, which is more than two centuries old. America thinks in feet; Europe thinks in meters. This causes different viewpoints during the standardization development.

Hierarchy of Standards

Every standard has its typical level of detail that must be agreed upon before work starts. Usually, the IT domain recognizes three levels: abstract, implementation, and interface. Standards at the abstract level are independent of operating systems, applications, hardware, and encodings. Standards at the implementation level determine encoding. Standards at the interface level determine hardware or hardware-oriented software, often called firmware. As a general principle, ISO standards reside at the abstract level. Most industry standards are at the implementation level or the interface level.

15.2.2 International Standardization Organizations and Consortia

An individual trying to find the standard that is relevant to an application may encounter a large number of groups and organizations that claim to be competent in the standardization business. At the international level, ISO and OGC are well known to the geographic information community. However, other acronyms such as IEC and ITU (defined below) are less familiar. In addition, the standards community works with a lot of abbreviations, such as TC, SC, JTC, WG, and SIG (defined below), which makes them difficult to read and understand. Nevertheless, once the meaning of the abbreviations is known, they can simplify communication. This section reviews the important groups, explains their terminology, and describes their interrelations.

The groups can be subdivided into two categories: *international organizations* and *international consortia*.

International organizations base their decisions on consensus. By having a budget scheme that allocates the financial burden of the organization to all member countries according to their economic potential, international organizations are fairly independent of the interests of individual governments or industries. Most of the organizations have a long history—some of them are almost 150 years old—although today their work rate is sometimes considered to be too slow for industry needs.

Three international organizations dominate the field of standardization. They are sometimes called the standards developing organizations (SDO): the ISO is the International Organization for Standardization, the IEC is the International Electrotechnical Commission, and the ITU is the International Telecommunication Union.

The members of *international consortia* are drawn primarily from industry, often from government agencies, and universities are occasionally represented. The primary goal of international consortia is to bundle the interests of their members. One of their interests is the development of common standards to advance other developments. Though the standard development is done with the participation of all members, the strong influence of the larger companies cannot be neglected. Though the standards might not be built on a broad consensus, the results are generally technically feasible and foster progress on the subject.

The OGC could currently be considered the most important consortium in the geographic information community. Another example of a consortium is the World Wide Web Consortium (W3C). It organizes the work necessary for the development and evolution of a Web technology into activities.

15.2.3 Formal International Standardization Organizations

Originally, the three organizations (ISO, IEC, and ITU) had their well-defined fields of activity. The IEC dealt with electrotechnical equipment, the ITU dealt with the radio transmission, and the ISO was responsible for all other subjects. However, today's sectors such as computer science have developed, and the borders between the subjects have become blurred. There are current demands to merge the work of all three organizations into one enterprise.

International Organization for Standardization (ISO)

The International Federation of the National Standardizing Associations (ISA) was founded in 1926, with its primary focus on mechanical engineering. The ISA ceased activities in 1942, owing to World War II. In 1947, the ISO was established as a new nongovernmental organization and continued the work. The ISO Central Secretariat is in Geneva, Switzerland.

People are sometimes confused by the mismatch between the name International Organization for Standardization and the three letters ISO. In fact, ISO is not an acronym, but rather a word derived from the Greek *isos*, meaning *equal*, which points to one of the goals of international standardization. This name is used around the world to denote the organization, thus avoiding the plethora of acronyms resulting from the translation of the full name into many different languages.

The work of ISO is based on three principles:

1. Consensus
   The views of all interested parties are considered. This is sometimes referred to as the democracy within the standardization development. In practice, influence in shaping a technical standard is largely restricted to the parties that can afford to pay for their experts to be involved. However, the final vote on a draft for an international standard depends on the agreement of 75% of the voting parties, independent of their activities during the development process.
2. Industry-wide
   The standards will always lead to global solutions that satisfy the needs of industries and customers worldwide. Due to diverse developments in different parts of the world, this principle often leads to a minimum consensus and to abstract-level standards. If worldwide consensus on a specific subject turns out to be impossible, the ISO would then withdraw its involvement.
3. Voluntary
   International standardization is market driven and therefore based on voluntary involvement of all interested parties. If a technical subject requires standardization, the interested parties would approach ISO, asking for guidance on the standardization process. ISO will guarantee compliance with the consensus principle and consistency with other international standards. The development will then receive recognition as an international ISO standard.

Members of ISO

ISO is the umbrella organization for national standardization activities. ISO members are primarily countries, which are usually represented by their respective national standards organizations. Examples of such national bodies are the American National Standards Institute (ANSI), the Standardisation Council of Canada (SCC), and the Deutsches Institut für Normung (DIN).

The national bodies are often unable to provide the complete range of expertise required for standard setting. Therefore, scientists and engineers usually exchange their knowledge within dedicated international organizations such as the International Cartographic Association (ICA) or the International Society for Photogrammetry and Remote Sensing (ISPRS). Companies express their interests through consortia such as the OGC. To incorporate this expertise, ISO has created another membership type called the external liaison organization.

In an effort to maintain consistent standards across committee borders and to avoid duplication of work, formal relations are established among ISO technical committees or subcommittees dealing with similar subjects. These relationships are called internal liaisons.

Technical Committees (TCs) and Subcommittees (SCs)

The work of ISO is decentralized. While the ISO Central Secretariat is mainly concerned with the development of consistent standards according to the ISO regulations, the actual work is done in the technical committees (TC) and the subcommittees (SC). Every TC and SC has its own secretariat at a competent institution somewhere in the world; for example, the secretariat for ISO/TC 211 *Geographic information/Geomatics* is attached to the Ordnance Survey, UK, in Southampton. The chairman has a powerful position within the TC, because he decides on the direction of the TC.

The number of members of a TC may range from 10 to 1000. In the case of TC 211, there are presently (2021) about 200 members, who meet about twice a year. In addition, today's communication technology allows almost daily exchange of notices and documents.

A technical committee is subdivided into working groups (WGs) that are usually responsible for the development of an individual standard or similar deliverable of the ISO. If the subject is very broad, the WG may be subdivided into two or more project teams that then become responsible for a single standard; for example, the standard ISO 19115-1 *Metadata – Part 1: Fundamentals* has been developed by a project team within the ISO/TC 211 WG 7 Information communities.

If a TC becomes very large, it has been found to be advisable to place the work of chairing the committee on more than one pair of shoulders. It is for this reason that the ISO established subcommittees (SCs). A TC may have many, one, or no SCs. It is a matter of power and politics whether a SC is

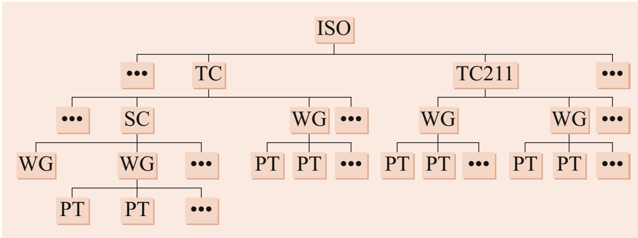


Fig. 15.1 Structure of ISO committees: TC, SC, WG, and project team (PT)

established or not, because the chairman has essentially the same influence as the chairman of a TC.

The TCs and the SCs are standing committees, which enables them to provide the maintenance that a standard requires in a long-term perspective. In contrast, the project teams and working groups are dissolved once a standardization project has matured. The TC and the SC initiate a review of each standard after, at most, a 5 year interval. The outcome of this review may be confirmation, revision, or withdrawal of the standard.

The long-term cooperation of the TC members fosters personal contacts and often leads to new technical ideas. A TC is always open to new members and expertise and can be compared to a large family. At present, ISO has 256 technical and project committees, with some of them being subdivided in subcommittees (end of 2021) (Fig. 15.1).

The ISO/TC 211 has no subcommittees (SC). Other technical committees such as ISO/TC 20 *Aircraft and space vehicles* have some subcommittees that function as another structural level. The technical committees that have been installed more recently tend to avoid subcommittees as this additional structural level.

Standardization Procedure in ISO

The idea for a new standard is always born outside ISO, and the ISO's role is to formalize a method to forward the idea to an internationally approved standard. The steward of the work is a technical committee or a subcommittee. The development process for a standard is subdivided into six consecutive stages:

1. *New work item proposal (NP)*. If an industry sector identifies the need for a new standard, then the industry representatives will ask their national standardization organization to send a new work item proposal (NWIP) to ISO. The ISO Central Secretariat assigns this new work item proposal to the most relevant technical committee (TC) or subcommittee (SC) and sends it out for a vote. Only if the majority of the P-members (participating members) of the TC or the SC vote in favor of the new work item proposal and at least five national bodies agree to actively participate in the working group will ISO advance the process. During this stage, the project team is formed and the members reach a consensus about the objectives of the planned standard.
2. *Building expert consensus*. The TC or SC sets up a working group and identifies a chairman referred to as the convener. The working group is asked to prepare a working draft (WD) as a first written summary of the future standard. Successive working drafts may be considered until the working group is satisfied that it has developed the best technical solution to the problem being addressed. At this point, the draft is forwarded to the working group's parent committee for the consensus-building phase. During this stage, the first version of the standard is developed.
3. *Consensus Building within TC/SC*. As soon as a first committee draft (CD) is available, the ISO Central Secretariat registers the document. It is distributed to all members of the TC or SC for comments. Within this stage, the document is reviewed carefully by independent experts who have not taken part in the internal discussions. During this stage, the standard reaches maturity.
4. *Enquiry on DIS (Draft International Standard)*. In this stage, the document becomes a DIS. Over a period of 5 months, the ISO Central Secretariat circulates the document to all ISO member bodies for voting and comments. It is only approved for submission as a final draft international standard (FDIS) if a two-thirds majority of the P-members of the TC or SC are in favor, and not more than one-quarter of the total number of votes cast are negative. If the approval criteria are not met, the text is returned to the originating TC or SC for further study. During this stage, the standard is aligned with the existing standards of ISO and with other ongoing work.
5. *Formal vote on FDIS (Final DIS)*. The ISO Central Secretariat circulates the FDIS to all ISO member bodies and requests that a final yes/no vote be made within a period of 2 months. Technical comments that are received during this period are not considered but will be registered for consideration during future revisions of the international standard. The text is approved as an international standard if a two-thirds majority of the P-members of the TC or SC are in favor and not more than one-quarter of the total number of the votes cast are negative. If these approval criteria are not achieved, the standard is referred back to the originating TC or SC for consideration in the light of the technical reasons submitted in support of the negative votes cast. During this stage, the standard has fulfilled all formal ISO requirements. In order to meet requirements for acceleration of the formal standardization procedure, the ISO Central Secretariat has started to weaken the rules and eventually publishes a positively voted DIS as an international standard, thus skipping the approval stage, if major comments are not anticipated.

6. *Publication of International Standards*. Once an FDIS has been approved, only minor editorial changes are made to the final text. The ISO Central Secretariat publishes the international standard (IS).

When difficulties are encountered in achieving consensus, the standardization process may take up to 5 years. Many standards are completed just prior to this deadline in order to avoid all the efforts of the team being officially deleted. If the parties still require the standard after the 5 year deadline has passed, then the whole process has to start over again.

Deliverables of ISO

The ISO is commonly considered an organization that makes standards or, more precisely, leads to their development and finally publishes them as binding international standards (IS). However, this is only partially true. Firstly, until around 1970, ISO only published international recommendations. Secondly, ISO has defined a reservoir of tools with simpler development procedures than full international standards. These other tools may be taken into consideration if a full international standard is out of scope as a result of slowness, lack of consensus, or simply a different intention such as informative reporting.

The products of ISO are standard documents. They are generally called deliverables. The following list shows all deliverables:

- ISO Standard
- ISO/TS – ISO Technical Specification
- ISO/TR – ISO Technical Report
- ISO/PAS – ISO Publicly Available Specification
- IWA – International Workshop Agreement
- ISO Guide.

ISO Standard

An ISO standard is normative. It is assumed that stages 1–3 of the standard development process are completed. In stage 4, the enquiry on DIS, the ISO Central Secretariat sends out the DIS to all ISO members for vote within a 5 month period. If two-thirds of the voting members (P-members) vote in favor and no more than one-quarter vote against, the DIS is approved. It moves onto stage 5, the formal vote on FDIS. After integrating comments, the document is issued as a FDIS for a final vote with the same voting procedure as with the DIS, but with a period of only 2 months. If approved, it is published as an IS, with a review of the IS taking place at least every 5 years.

ISO/TS – ISO Technical Specification

An ISO technical specification (ISO/TS) is normative as well. After completion of stage 3 of the standard development, the committee stage, the document is handed to the TC or the SC. This document is called a committee draft (CD). The TC or SC sends out the CD for vote within a default period of 8 weeks. If two-thirds of the P-members vote in favor, the CD is approved as an ISO/TS. An ISO/TS must be reviewed after 3 years. After 6 years, it must be advanced to an international standard or withdrawn.

An ISO/TS is a means to make a document official if there is insufficient support for an international standard, or if other existing specifications will get the ISO brand.

An example of an ISO technical specification is the ISO 19103 *Conceptual schema language*.

ISO/TR – ISO Technical Report

An ISO technical report (ISO/TR) is informative. At any stage after the approval of an NWIP the ISO Central Secretariat may decide to publish a document as an ISO/TR. The ISO/TR requires a simple majority of the P-members of the TC or SC for approval.

ISO/PAS – ISO Publicly Available Specification

An ISO publicly available specification (ISO/PAS) is also normative. After completion of stage 2 of the standard development, the preparatory stage, the working group must find consensus on the document. A simple majority of the P-members of the TC or SC is then necessary to create an ISO/PAS. After 6 years, it must be advanced to an international standard or withdrawn.

None of the few examples of ISO/PAS documents is relevant to geographic information.

IWA – International Workshop Agreement

An international workshop agreement (IWA) is normative. It is an ISO document that is prepared outside the formal structures of ISO, such as the TCs. The market players, mostly industries, instead agree upon the IWA in an open workshop environment in fields where ISO has no experts or structures available. The IWA is one of ISO's strategies to accelerate its response to market requirements. The option exists to develop the IWA into a full international standard at a later stage.

ISO Guide

ISO guides provide guidance to technical committees for the preparation of standards, often on broad fields or topics.

International Electrotechnical Commission (IEC)

The IEC is the global organization that prepares and publishes international standards for all electrical, electronic, and -related technologies. The IEC was founded in 1906. It is a nongovernmental organization with its central office in Geneva, Switzerland. At present, the IEC has 62 member countries and 24 associated member countries.

IEC and ISO follow almost identical standardization procedures, with standards built on consensus among the interested parties, worldwide solutions, and voluntary initiatives.

In general, the IEC uses the same terms for the six stages leading to an international standard: the proposal stage, the preparatory stage, the committee stage, the enquiry stage, the approval stage, and the publication stage.

The ISO/IEC Joint Technical Committee 1 (JTC1)

The ISO and the IEC were founded by the mechanical engineering and electrical engineering communities, respectively. Although the separation of the two organizations might have been logical at one time, the era of computers required joint efforts involving both organizations. In 1987, the Joint Technical Committee 1 (JTC1) was created to provide a single, comprehensive standardization committee that addressed international information technology standardization.

The name is confusing in that it implies the existence of other JTCs like JTC2 and JTC3. However, the only joint technical committee in existence is the JTC1.

The International Telecommunication Union (ITU)

Founded in 1866, the ITU administers the worldwide use of radio frequencies and satellite orbits. It is the oldest of the three large standards developing organizations. As the ITU is an international organization within the United Nations (UN) system, it represents all countries in the world, presently numbering 193. The Secretary General is located in Geneva, Switzerland.

The ITU is subdivided into three sectors: the radiocommunication sector (ITU-R), the telecommunication standardization sector (ITU-T), and the telecommunication development sector (ITU-D).

The activities of the telecommunication standardization sector mainly lead to standardization for network hardware and network protocols. The work is distributed to a number of study groups. The deliverables are recommended standards for international use. As the UN is a treaty organization, the recommended standards immediately become national laws in many countries.

15.3 Geomatics Standards

The reader will be familiar with GIS and will think in terms of the GIS with which he or she has gained experience. Some might think in terms of digital maps that are available on the Internet and others in terms of their areas of responsibility: property cadastre, environmental applications, or fleet management systems. Another important application might be disaster management, where a GIS can help save lives if rescuers receive an almost immediate and detailed picture of the environment of the disaster location.

This section deals with the standardization of geomatics. Worldwide standardization of geomatics is not simple, because an enormous number of entirely different applications are served by GIS techniques. For instance, applications can range from city maps delivered on a CD to global climate monitoring. Geomatics applications run on a single workstation or access thousands of computers connected in a network. Overall, a GIS consists of a number of rather independent modules serving data capture, data storage, or data exchange, to mention a few.

The reader might think that some kind of typical GIS should become an international standard. This would make the work easier, because all *standard GIS* would have a similar shape in terms of their core components. Most users demand standardized data exchange formats, because problems with data transfer are well known.

ISO standards are developed with a long-term perspective. Most of them are written at an abstract level to guarantee long-term stability. In contrast, developments in the geomatics domain continue to make fast progress. Some of the ISO standards for geomatics are implementation oriented, even though this is not the first priority of the ISO/TC 211 work.

15.3.1 ISO/TC 211

History and Work of ISO/TC 211

Europe has long-established traditions in the development of standards for cadastre, cartography, and environmental protection. Following these traditions, the first project to standardize geographic information started in Europe in 1991. Led by the Association Française de Normalisation (AFNOR), the national standards body of France, the European standardization organization CEN created its Technical Committee 287 *Geographic information*. The work of CEN/TC 287 resulted in eight European prestandards. The successful conceptual developments were integrated into the larger body of ISO a few years later. The work of CEN/TC 287 originally contained a full range of nearly 20 standards. Facing the upcoming establishment of ISO/TC 211 in 1994, and to avoid duplicate work, CEN/TC 287 ceased its activities and went into a dormant state until 2003.

The driving forces behind the establishment of ISO/TC 211 were the North Atlantic Treaty Organization (NATO) Defence Geospatial Information Working Group (DGIWG) and the national standards efforts in the USA and Canada. The other two organizations that contributed experiences in the standardization of geographic information were the International Hydrographic Organization (IHO) and CEN/TC 278 *Road transport and traffic telematics*. CEN/TC 287 *Geographic information* had an established program of work that essentially became the plan for the ISO/TC 211 base standards. It was the DGIWG that originally proposed the formation of ISO/TC 211, and because it was procedurally easier to have a nation make the proposal, Canada made the proposal in 1994.

Both the original CEN work and the DGIWG work were closer to the implementation level than the current ISO/TC 211 standards. Over time, the ISO standards have become more abstract base standards that require profiles and implementation specifications in order to be put to use.

With the establishment of ISO/TC 211, a joint worldwide effort to standardize geomatics began. In particular, ISO/TC 211 integrated the European and North American experiences with other regions of the world, including Asia, Australia, and South Africa, which have since joined the committee.

Scope

The objective of the work of ISO/TC 211 is to establish a set of standards for geographic information/geomatics. The standards would specify an infrastructure and the required services for handling geographic data, including management, acquisition, processing, analysis, access, presentation, and transfer. Where possible, the standards would link to other appropriate standards for information technology and provide a framework for the development of sector-specific applications [15].

Structure

An ISO technical committee consists of the chairman, the members, the working groups, some advisory groups, and, eventually, some subcommittees.

The Chair, elected by the members of ISO/TC 211, is Peter Parslow from the Ordnance Survey, UK. He followed Agneta Engberg from the Swedish national and cadastral agency, "Landmäteriet", who served as chair from 2016 to 2021.

The members are national members and liaison members. The *national members* are represented by the authorized standardization organizations of the respective member country, also called the *national body*. Normally, the national bodies are *participating members* (P-members) with full voting rights, while others are *observing members* (O-members) with observer (nonvoting) status only. The liaison organizations are domain specific, worldwide or regional organizations that can contribute expertise to the standardization process. Examples of non-ISO organizations (referred to as *external liaison members*) are the Open Geospatial Consortium and the International Hydrographic Organization. *Internal liaison members* are other ISO or IEC committees, such as ISO/TC 204 *Intelligent transport systems*, that also have interests in geographic information.

Representatives from liaison organizations are invited to take part in the discussions and can receive all information on the standardization project but are not granted voting status.

The working groups (WGs) form part of the ISO/TC 211 hierarchical structure, and as the name implies, they are the place where the standards are actually drafted. Until 2001, ISO/TC 211 had WGs 1–5; after the work of WGs 2, 3, and 5 finished in 2001, WGs 1 and 4 remained, and WGs 6, 7, 9, 10, and 11 were created. The WGs combine a number of individual standardization projects related to a common topic, such as WG 4 related to Geospatial services.

The technical committee creates advisory groups for special purposes, such as the advisory groups on strategy and on outreach. Such groups analyze the current and future requirements for geomatics standardization and draft proposals to the technical committees on how to proceed. The advisory groups do not belong to the working groups; they are directly responsible to the technical committee and support the chairman. The Harmonized Model Maintenance Group (HMMG) is working on the integration of all partly incompatible models of the ISO 19100 standards. The XML Maintenance Group (XML MG) supports the project teams in using the XML consistently when drafting the standards documents.

ISO/TC 211 has no subcommittees (as of November 2021).

Statistics

The work of ISO/TC 211 can be described in the following numbers (November 2021):

- ISO deliverables
  - About 85 international standards (IS) completed
  - Presently about 20 projects active
  - 1st international standard: ISO 19105:2000 *Conformance and testing*
  - 1st technical report: ISO/TR 19121:2000 *Imagery and gridded data*.
- Members
  - 37 P-members
  - 33 O-members
  - 39 external liaison members
  - 32 internal liaison members
  - relation to CEN technical committees (Europe)
  - 11 advisory groups and several other groups.
- Meetings
  - 53 plenary meetings.

Program of Work

In 1994, ISO/TC 211 started with only 20 standardization projects. These formed a suite of base standards for geographic information and, thus, made the technical committee a coherent and powerful group. Some experts consider these abstract base standards as the original obligation of ISO/TC 211. The base standards included the reference model, feature definition, spatial and temporal schema, co-

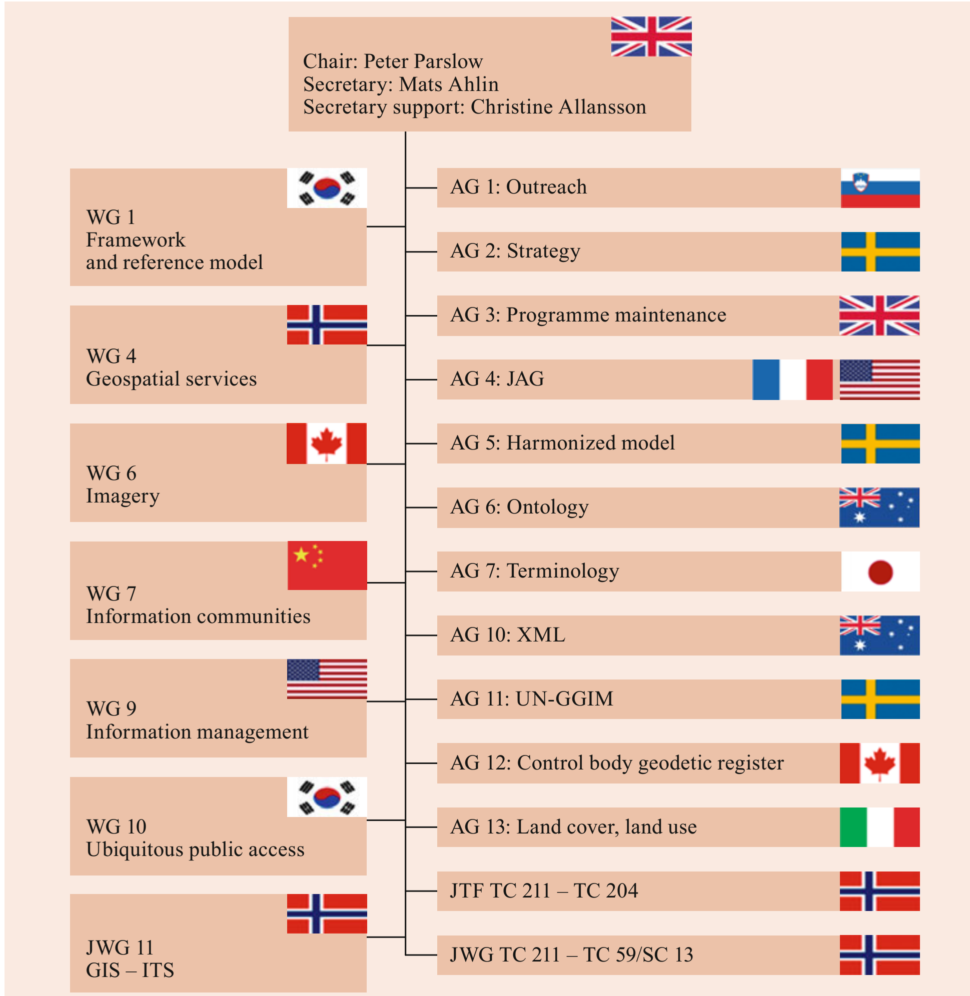


Fig. 15.2 Structure of ISO/TC 211 (status 2021): AG = advisory group, JAG = TC 211/OGC joint advisory group, JTF = joint task force, JWG = joint working group, WG working group, ITS = Intelligent Transport Systems.

ordinate reference system, portrayal, encoding, quality, and metadata, to mention the most important ones.

ISO/TC 211 is presently undergoing significant changes in order to meet the challenges of the future (Fig. 15.2):

1. Since the completion of the base standards, interest has been focused on their implementation in real-world systems. Admittedly, implementation issues are outside the scope of ISO, but they are the only way of proving the viability of the original abstract development. It has been recognized that the field of geographic information is so heterogeneous that the base standards are not always able to fully meet the requirement for a generic and consistent approach that is expected from an ISO suite of standards. Their implementation produces the necessary feedback to optimize the base standards. The cooperation of the OGC is of particular importance for implementation issues, because some of their implementation-type standards, such as GML, have become ISO standards.
2. Emerging technologies have started to widen the original scope of ISO/TC 211. The industry now expects a tremendous growth in the market for location-based mobile services, as well as four-dimensional data and prediction modeling in the near future, and is pushing for standardization at the boundaries of geographic information, general databases, and mobile positioning systems.
3. The base standards only focused on vector geometries, as the field of gridded data – digital images, coverages, and elevation models – seemed too large and unknown to be included in the first attempt at geomatics standards. The projects on imagery standards started in 2001.

Example of an ISO-Compliant GIS

This section provides a basic understanding of the interrelation between the standards of the ISO 19100 family. This understanding is communicated via a small tourism information GIS. This GIS is decomposed from the GIS professional's perspective in order to relate every component to an individual ISO 19100 standard and to explain which details are standardized and which details are outside the scope of ISO 19100.

Fig. 15.3 Example of a tourism GIS

Characterization of the Example GIS

The example is a tourism GIS, as shown on the map in Fig. 15.3. It contains topographic data, such as the lake, the woods, and the roads. It also contains thematic data, such as the delineated trails as hiking trails and the opening hours of the theater. The map is a partial representation of the data and their structure. The map shows only the graphics. The other data might be shown by another type of map or by an interactive program on the computer screen.

The following paragraphs summarize other information that belongs to the GIS *alphanumerically*. These data are typical for this kind of GIS.

A company called Tourism-GIS-Association may be responsible for any aspect related to the example GIS.

The GIS data, such as the topographic base map and the information center for the local community, have different origins; for instance, the geometry, including information content, may have been copied from a topographical base map. The state mapping authority provided additional information about the last revision date and the geometric accuracy of the data. The local community may have supplied the delineation of some of the trails as official hiking trails and the opening times of the campground and the theater.

From the GIS professional's point of view, the GIS may be decomposed into the components of data capture, data storage, and data display. In the tourism GIS, the data capture has been completed by introducing the various components from other resources, as mentioned above. We can assume that no extra data were surveyed in the field for use in the GIS.

A database will be used for data storage. The data are not simply lines as shown on the map, but rather are grouped into so-called objects, such as the theater. Additional data, such as the opening hours, can be linked to this object. Another object is the lake, including its shoreline. The objects themselves consist of points, curves, and surfaces and are defined by their coordinates. The mosaic of all objects with the graphic type surface is equal to the complete area covered by the GIS.

A data-editing component of the tourism GIS could include a user-accessible function to compute, for example, the approximate lengths of the hiking trails. The user may have to click on the graphics of the trail to cause the system to display the trail's length.

The functionality of the data display component is the shown map. In a general sense, the display is controlled by a reference table that relates every type of object to a graphic representation, such as two parallel lines for roads.

Relation to the ISO 19100 Standards

The ISO 19100 family of standards does not form a complete and hierarchical model for the whole universe of geographic information. The standards are more a collection of independent abstract standards for creating and managing GIS. Most standards are only loosely linked.

In the terminology of the ISO 19100 standards, the tourism GIS is an application schema. ISO 19101-1 *Reference model – Part 1: Fundamentals* defines the components of the application schema, as shown in Fig. 15.4.

Data Capture

ISO does not standardize the data capture procedures for a GIS. The ISO 19100 standards only provide guidelines and metadata elements to describe the origin and quality of the data. The relevant standards are ISO 19157 *Data quality* and ISO 19115-1 *Metadata – Part 1: Fundamentals*.

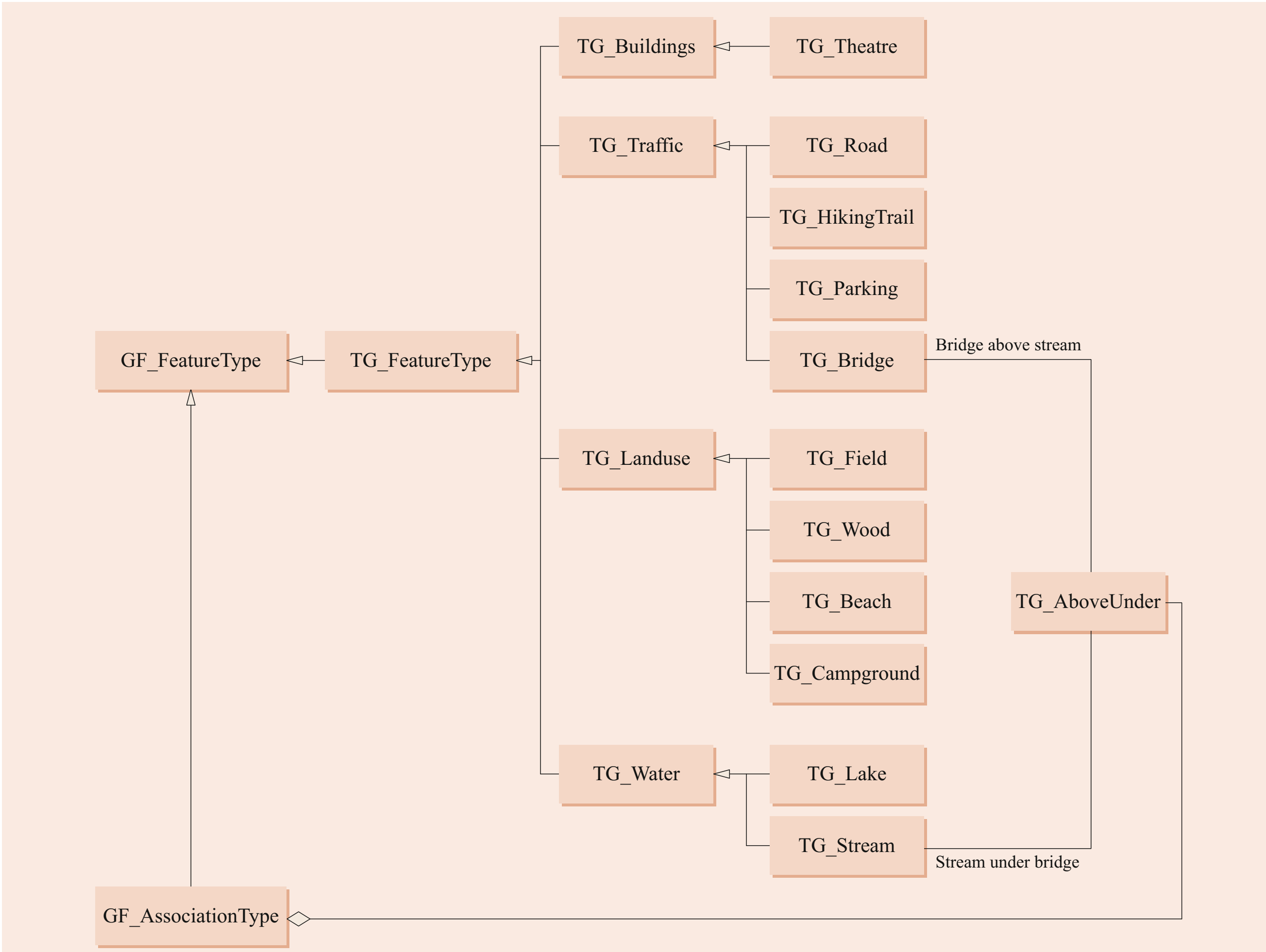


Fig. 15.4 Data model of the tourism GIS shown as a UML class diagram. All classes of the tourism GIS start with the two characters "TG" to indicate their relation to the tourism GIS. UML is addressed in Chap. 1

The following paragraph provides a detailed relation of the tourism GIS to the quality concept of the ISO 19100 standards. Let us assume that we need to write a *quality evaluation report* about our GIS. This report is a standardized procedure of ISO 19157. It might report values of the *data quality elements*, such as:

- Completeness: 100% coverage
- Logical consistency: no errors found
- Positional accuracy: $\pm$ 5 m
- Temporal accuracy: last map revision 3 years ago; community information supplied last spring
- Thematic accuracy: topographic base map does not supply shelters and other similar elements for tourist purposes.

It might also report values of the data quality overview elements, including:

- Purpose: tourism GIS will enable an online information source of tourist details, and it will also promote tourism in the area.
- Usage: dataset was assembled for the tourism GIS.
- Lineage: dataset was created under the supervision of the Tourism-GIS-Association.

ISO 19157 *Data quality* provides advice on different levels of detail for the quality check.

Let us assume that we have only checked the purpose, usage, and lineage, and that we will write the results in a report. ISO 19115-1 *Metadata – Part 1: Fundamentals* provides the formal names of all the necessary data elements.

Conceptually, ISO 19115-1 covers the complete list of metadata elements supporting the description of all aspects of the tourism GIS data: from its producer, date of production, and cost, to format, where it can be obtained, and many other item of information about the data. A part 2, ISO

Table 15.1 Feature catalogue of a tourism GIS

GF_FeatureType high level	GF_FeatureType low level	GF_Operation	GF_AttributeType	GF_AssociationType
Buildings	Theater	Closed in winter	Point	
Traffic	Road		Line	
Traffic	Hiking trail		Line	
Traffic	Parking		Polygon	
Traffic	Bridge		Point	Bridge over stream
Land use	Field		Polygon	
Land use	Wood		Polygon	
Land use	Beach		Polygon	
Land use	Campground	Closed in winter	Polygon	
Water	Lake		Spline	
Water	Stream		Spline	Stream under bridge

19115-2 *Metadata – Part 2: Extensions for acquisition and processing*, integrates further elements related to the development of the data.

Data Storage

In the terminology of the ISO 19100 standards, the tourism GIS is referred to as a *dataset*. The elements of the dataset were called *objects* in the beginning of this section. However, although the term "object" is familiar in the context of object-oriented languages, in the domain of geographic information, the term "feature" is more widely used to denote an element of a dataset. The term "feature" has also been adopted by the ISO 19100 standards.

ISO 19109 *Rules for application schema* contains all the definitions related to features [16]. The core of ISO 19109 is the general feature model (GFM). It states that a feature may have attributes and operations. Attributes identify whether a feature is a point, a curve, or a surface, and operations identify whether the feature changes according to external influences, such as the theater being closed in winter. Features may be logically grouped into larger units; for example, woods and fields both belong to land use. A single element in the dataset, such as one of the hiking trails, is called an instance of the feature *hiking trails*. Some features have special relationships that become important once a map is drawn, or networking algorithms are applied to the dataset. An example for this case is the bridge crossing the stream. The bridge will always go over the stream at a different elevation above the water surface. This relationship is an association between features.

It is obvious that the possible features of a GIS must be agreed upon before population of the database starts. The result of this design process is basically a listing of all allowed features together with their attributes, operations, and associations. The listing is called a feature catalogue. The complete result is called a *data model*.

The ISO does not standardize feature catalogues and data models. The only guidance provided is the ISO 19110 *Methodology for feature cataloguing*. This standard provides the general rules and a catalogue template to support a complete and consistent listing [17].

The following list is a simplified feature catalogue for the tourism GIS (Table 15.1).

It was stated earlier that the attributes of a feature could be of type point, line, or area. In fact, this was an oversimplification. The ISO 19100 standards provide a great variety of geometry classes and rules for how they are related to each other. The most comprehensive collection of geometry classes can be found in ISO 19107 *Spatial schema*.

The tourism GIS uses four geometry classes of ISO 19107: Point, Spline, Line, and Polygon. The class Polygon denotes surface-type features.

ISO 19107 distinguishes between primitives and complexes. Primitives are geometries that *do not* include their end-points. Complexes are geometries that *do* include their end-points. If a topological network underlies our tourism GIS, the geometries must be connected at their end-points. Thus, the geometry consists of complexes.

The ISO 19100 family does not standardize the way the geometry and topology are handled by the database. It does not standardize data types or details of the topology. Advanced systems only store one point at one position, allowing rapid searching of neighboring features. Simpler systems only store individual geometries and allow multiple points at a position.

All geographic positions of the data are defined according to a coordinate reference system. ISO 19111 *Referencing by coordinates* sets the rules for the definition of coordinate reference systems. In the case of our tourism GIS, a local coordinate reference system is a sufficient solution. A local coordinate reference system is an *engineering* coordinate reference system according to ISO 19111. To simplify the usage of coordinates it is advisable to avoid negative values. For this purpose, the complete definition area of the GIS must lie in the first quadrant, or seen from another perspective, the origin of the coordinate reference system must be beyond the west and south of the definition area. The x and y-axis might point east and north, respectively. The datum (= origin) and

Table 15.2 Portrayal catalogue of a tourism GIS

PF_FeaturePortrayal	PF_PortrayalRule
Beach	Light brown area, border with black solid line, thickness 0.2 mm, letter B at the representativePoint
Bridge	Place bridge symbol
Campground	Light brown area, border with black solid line, thickness 0.2 mm, letter C at the representativePoint
Field	Light brown area, border with black solid line, thickness 0.2 mm
Hiking trail	Dashed black line, line 2 mm, space 2 mm, thickness 0.2 mm
Lake	Light brown area, border with black solid line, thickness 0.2 mm
Parking	Light brown area, border with black solid line, thickness 0.2 mm, letter P at the representativePoint
Road	Two black solid parallel lines with thickness of 0.2 mm each, one line 1 mm to the left and the other line 1 mm to the right of the road's axis, interrupt lines to open roads at intersections, show roads at highest priority if overlay with other graphics occurs
Stream	Black solid line, thickness 0.2 mm
Theater	One circle, and two concentric semicircles, black solid lines, thickness 0.2 mm
Wood	Light brown area, border with black solid line, thickness 0.2 mm, place tree symbols at irregular positions

the prime meridian (= axes to north) provide a local reference only. ISO 19111 does not standardize any parameter describing global or local coordinate reference systems.

Descriptive parameters are kept independent of the ISO 19100 family of standards. To guarantee standardized use of important parameters, ISO 19100 provides one or more registries. The rules for creating these registries are described in ISO 19135-1 *Procedures for item registration – Part 1: Fundamentals*. A specific registry for coordinate reference systems is defined in ISO 19127 *Geodetic codes and parameters*. If parameters are needed, they can be requested from one of the ISO-approved registration authorities.

Data Display

According to ISO 19117 *Portrayal*, the graphic representation is handled independently from the feature data. In the case of the tourism GIS, an individual feature portrayal is defined for each of the features such as the roads and hiking trails. All necessary feature portrayals are summarized in a listing called the *feature catalogue*, which contains one entry for each feature in the dataset. Consequently, the feature catalogue of the tourism GIS has 11 entries.

Each feature portrayal points to a portrayal function that will be applied when drawing that feature. To handle more sophisticated graphics, such as the automatic opening of a road intersection, ISO 19117 allows the usage of external functions. The summary of all portrayal elements is the portrayal specification. During the output of the feature data, this specification is applied to the data stream to generate the cartographic representation.

Table 15.2 presents the portrayal catalogue of the tourist GIS (class PF PortrayalCatalogue).

Data Exchange

ISO does not standardize any exchange format, but data exchange is addressed in ISO 19118 Encoding, which recommends that an exchange format be built upon the currently widely used XML. Also, ISO does not deal with the technical details of transferring data between different hardware platforms. ISO 19118 only states that the application schemas of the two systems between which the data exchange takes place must be alike. This simply means that a system to which the tourism GIS data may be transferred has to provide the same feature types: lake, wood, beach, etc., and the same structural information. If the feature types do not match, information is lost during the transfer. If, for example, in the second system only one feature for traffic is available, then roads and hiking trails would become the same feature and could not be distinguished from one another.

15.3.2 Roadmap to the ISO 19100 Standards

Figure 15.5 shows a roadmap for all ISO 19100 standards.

The roadmap shows the structuring of ISO 19100 standards used in this chapter. The top-categories are not official terms. However, they are used in internal discussions, and they are helpful to understand the overall structure of the ISO 19100 family.

15.3.3 Reference

Standardization of geographic information covers a wide range of tasks, including the definition of interoperability, fundamental data types such as for spatial and temporal information, modeling rules, and the semantics of real world phenomena. A reference model is set in order to achieve these tasks in an integrated and consistent manner.

Reference Model – Part 1: Fundamentals (ISO 19101-1)

Among geomatics specialists, a consensus has emerged that the field of geographic information is a specialization within the information technology field [18]. At the same time, there is a growing recognition among users of information technology that indexing by location is a fundamental way

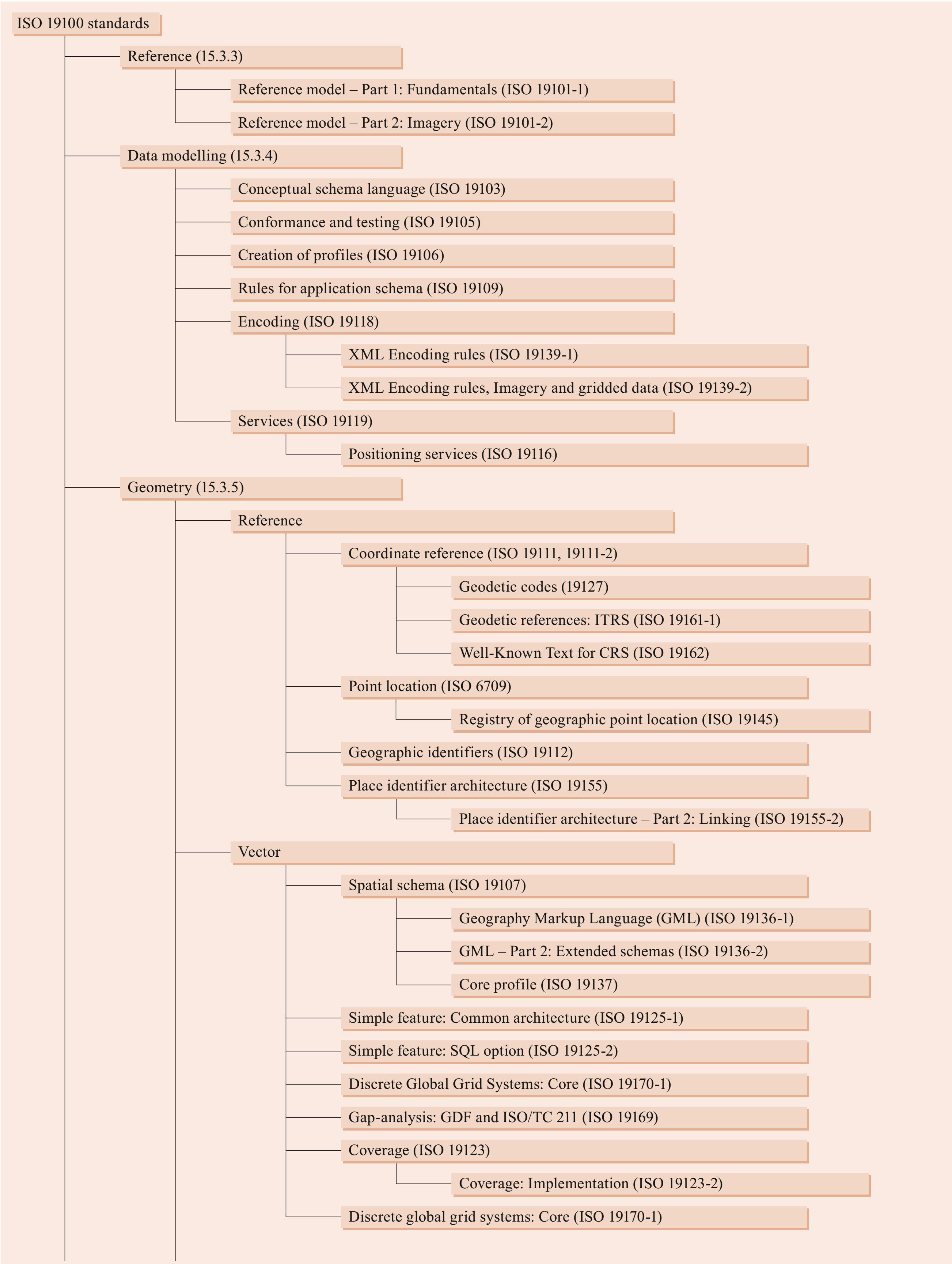


Fig. 15.5 Roadmap to the ISO 19100 standards

- Raster
  - Framework (ISO/TR 19121, ISO 19129)
  - Sensors: Fundamentals (ISO 19130-1)
  - Sensors: SAR, InSAR, Lidar and Sonar (ISO 19130-2)
  - Sensors: Implementation (ISO 19130-3)
  - Calibration and validation: Optical sensors (ISO/TS 19159-1)
  - Calibration and validation: Lidar (ISO/TS 19159-2)
  - Calibration and validation: SAR/InSAR (ISO/TS 19159-3)
  - Content components: Content model (ISO/TS 19163-1)
  - Content components: Implementation (ISO/TS 19163-2)
- Visualization: Portrayal (ISO 19117)

- Terminology (15.3.6)
  - Terminology (ISO 19104)
  - Cross-domain vocabularies (ISO 19146)
- Metadata (15.3.7)
  - Metadata: Fundamentals (ISO 19115-1)
  - Metadata: Extensions for acquisition and processing (ISO 19115-2)
  - Metadata: Implementation (ISO 19115-3)
    - Metadata – Implementation – Part 1: Encoding rules (ISO/TS 19139-1)
    - Metadata – Implementation – Part 2: Imagery (ISO/TS 19139-2)
  - Data product specifications (ISO 19131)
- Quality (15.3.8)
  - Data quality (ISO 19157)
    - Data quality – Part 2: Implementation (ISO/TS 19157-2)
    - Quality assurance (ISO/TS 19158)
- Time (15.3.9)
  - Temporal schema (ISO 19108)
- Catalogues (15.3.10)
  - Registration – Part 1: Fundamentals (ISO 19135-1)
  - Registration – Part 2: Implementation (ISO/TS 19135-2)
    - Feature dictionaries (ISO 19126)
  - Feature catalogues (ISO 19110)
- Web Services (15.3.11)
  - Web Map Server (ISO 19128)
  - Web Feature Service (ISO 19142)
  - Filter encoding (ISO 19143)
  - API for features: Core (ISO 19168-1)

Fig. 15.5 (Continued)

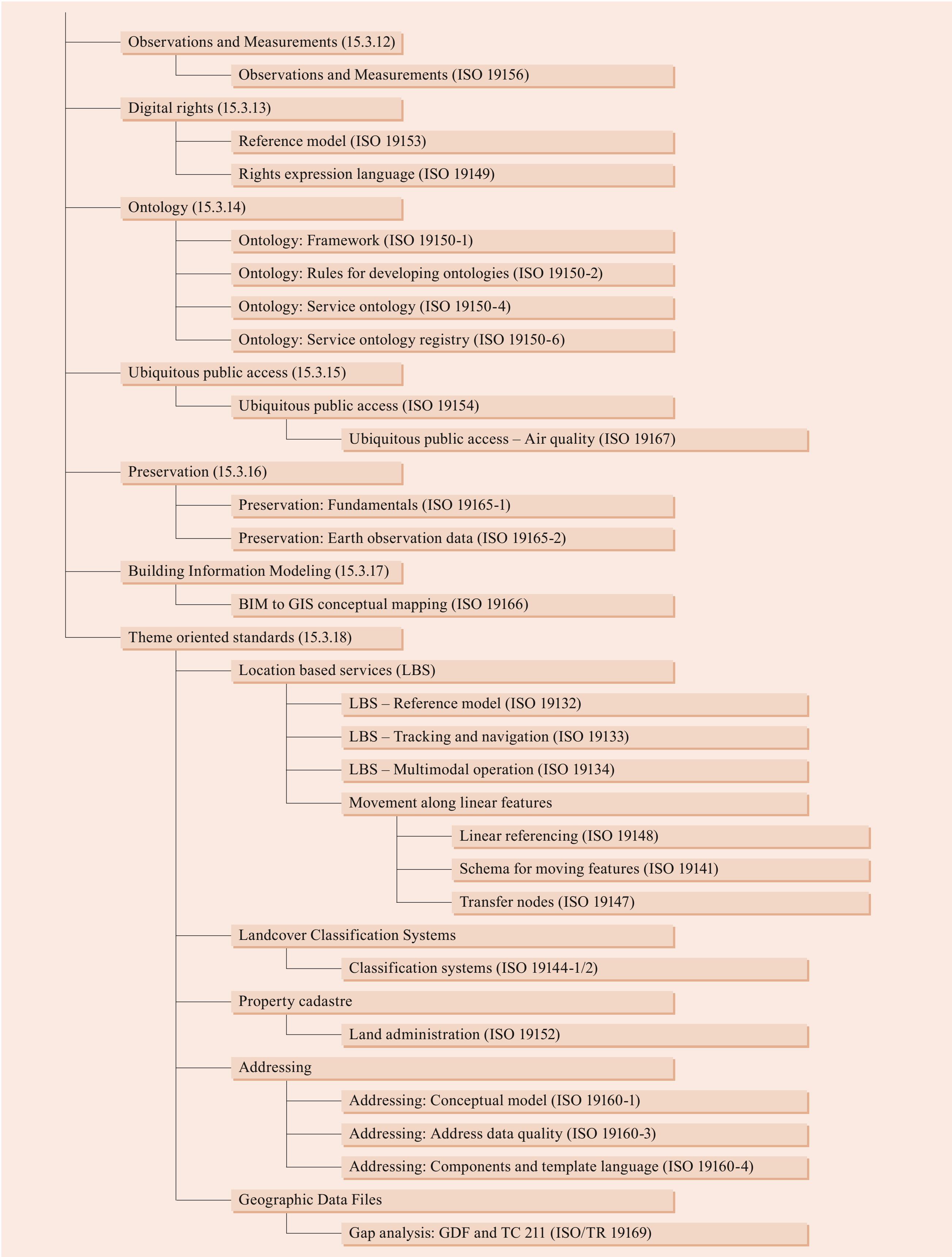


Fig. 15.5 (Continued)

of organizing and using digital data. To meet these needs, standardization of geographic information in the ISO 19100 series is based on the integration of the concepts of geographic information with those of information technology. Whenever possible, the development of standards for geographic information considers the adoption or adaptation of generic information technology. It is only when this cannot be done that geographic information standards have been developed. The basic principles of information technology can be found in standards of ISO and IEC.

Viewpoints

The usage of viewpoints is a common method of decomposing large, distributed software systems during the design process. This core model is standardized as ISO/IEC 10746 *Information technology – Open distributed processing – Reference model*. It is also the foundation of the reference model for the ISO 19100 family of standards.

ISO/IEC 10746 describes five viewpoints:

1. The *enterprise viewpoint* is concerned with the purpose, scope, and policies of an organization in relation to GIS. This viewpoint is used to generate requirements and varies among different organizations, and, therefore, it is not within the purview of the ISO 19100 family of standards.
2. The *information viewpoint* is concerned with the semantics of information and information processing. A specification developed from this viewpoint provides a model of the information in a GIS and defines the processing that is performed by such a system. The information provides a consistent common view on information that can be referenced in a GIS. The information viewpoint is the most important viewpoint for the ISO 19100 family of standards.
3. The *computational viewpoint* is concerned with the patterns of interaction among services that are part of a larger system. A service may be the interaction between the system and the client, such as the window interface on the screen, or between a set of other services, such as data retrieval from a database in the background. The computational viewpoint is the second most important viewpoint for the ISO 19100 family of standards.
4. The *engineering viewpoint* is concerned with the design of implementations within distributed, networked computing systems that support system distribution. As implementation is not the focus of the ISO 19100 standards, there is little emphasis placed on this viewpoint.
5. The *technology viewpoint* is concerned with the provision of the underlying hardware and software infrastructure within which services operate. Again, as this is out of scope of the ISO 19100 family of standards, there is little emphasis placed on this viewpoint.

Fig. 15.6 displays the viewpoints graphically.

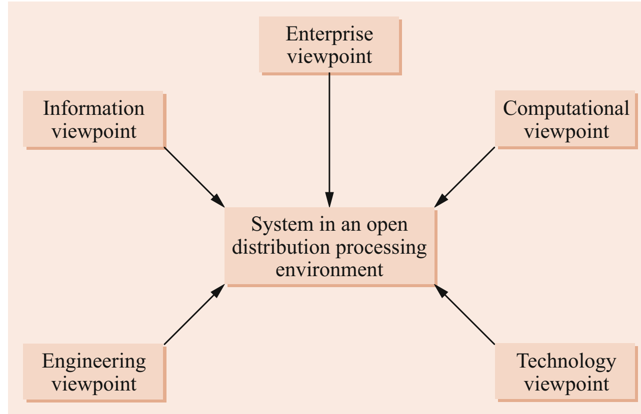


Fig. 15.6 Viewpoints of the ISO/IEC 10746 *Open distributed processing – Reference model*

ISO 19101-1 *Reference model – Part 1: Fundamentals* defines a hierarchically structured reference model for the ISO 19100 family of standards. Primarily, this reference model is a special application of the *information viewpoint*, which is identified as the most important one for the standardization of geomatics. One component of the reference model is based on the *computational viewpoint* that addresses the services.

The majority of individual standards of the 19100 family are considered as services that operate on datasets. The services are discussed in detail later.

ISO 19101-1 is primarily for standards developers. However, understanding this standard is also helpful for users of the other geomatics standards, because it leads to an understanding of the rationale of the ISO 19100 series.

Conceptual Schema Modeling Facilities

The background for structuring the reference model of geographic information is a number of principles that information technology has agreed upon. The withdrawn ISO/IEC 14481 *Conceptual schema modeling facilities (CSMF)* laid out a schema for the design of computer software that has four levels of abstraction.

The *meta-metamodel level* is the highest level. Though the name is somewhat obscure, the meaning of the meta-metamodel is simple. It is the description of a software system in natural language. The meta-metamodel level is not a matter of standardization.

The *metamodel level* is below the meta-metamodel level in the hierarchy. It contains the description of the concepts of a software system in a formalized language, which, in the case of the ISO 19100 standards, is the Unified Modeling Language (UML). The concepts include terminology, operations, and assumptions needed to construct applications. Typical examples of metamodels are the UML diagrams found in this handbook and throughout the ISO 19100 standards.

The *application model level* exists below the metamodel level. An application model contains all detailed definitions needed to tailor a software system to a specific application. In the case of geographic information, this might include the definition of feature types such as roads and buildings and their possible attributes and operations. The documentation of an application model is called an application schema.

The *data model level* contains the datasets and is the lowest level. In the case of geographic information, the datasets are composed of one or many features, including their attributes and positions. The datasets are the actual data, whereas the application schema only sets the frame for the possible features.

Metamodel

Any consideration on modeling starts with the real world. The chosen piece of the real world that one wishes to describe in a model is known as the universe of discourse. A first result of the modeling process is a conceptual model that describes and limits the universe of discourse. A model is an abstraction and represents only a part of the real world. An example of the real world could be the landscape depicted in the map in Fig. 15.3. A conceptual model is the list of features relevant for the tourism GIS in Table 15.1.

A conceptual schema language provides the semantic and syntactic elements used to rigorously describe the conceptual model and to convey consistent meaning. A conceptual model described using a conceptual schema language is called a conceptual schema. A conceptual schema of the tourism GIS is shown in Fig. 15.4. The conceptual schema language prescribed in the ISO 19100 series is UML. The feature types that are shown represent the semantics. The lines specifying the exact type of associations are the syntactic elements. The conceptual schema language is addressed in more detail later.

Application Model

The application model of a GIS completely defines its data elements (classes) and their interrelation. The heart of the application model is a detailed description of the data structures. This description is called an application schema. The tools to describe the data structure are the feature catalogue and the General Feature Model (GFM). The feature catalogue is a formal list of the feature types that are determined as relevant and qualified for the application. The GFM sets the frame for the definition of every feature type. An example of a feature catalogue is provided in Sect. 15.3.2. The GFM is explained in detail in Sect. 15.3.5. Both tools have dedicated ISO 19100 standards.

An application schema also consists of components that are relevant to all feature types. These components are the coordinate reference system and elements describing data quality. The coordinate reference system is explained in detail in Sect. 15.3.6, and the quality aspects are addressed in Sect. 15.3.9. Both components have their own standards in the ISO 19100 family.

Data Model

The data model contains the data being collected for an application and the services that operate on the data. The data model includes all features and their describing data that make up the complete GIS. The data consist of the dataset and the metadata. The dataset includes only the features and their positions.

The metadata is the describing data. ISO 19115-1 *Metadata – Part 1: Fundamentals* provides a good guideline for the selection of metadata. ISO 19115-1 provides a standardized list of metadata elements. Further describing elements that are not available in ISO 19115-1 may be modeled as metadata elements beyond the scope of ISO 19115-1 or as attributes of the feature classes.

Programs that use the dataset and the metadata to serve a certain purpose for the user are called geographic information services. In the ISO 19100 standards, it is stated that services operate on the dataset and reference additional information from the metadata. Most of the ISO 19100 standards define such services.

Conceptual Modeling and the Domain Reference Model

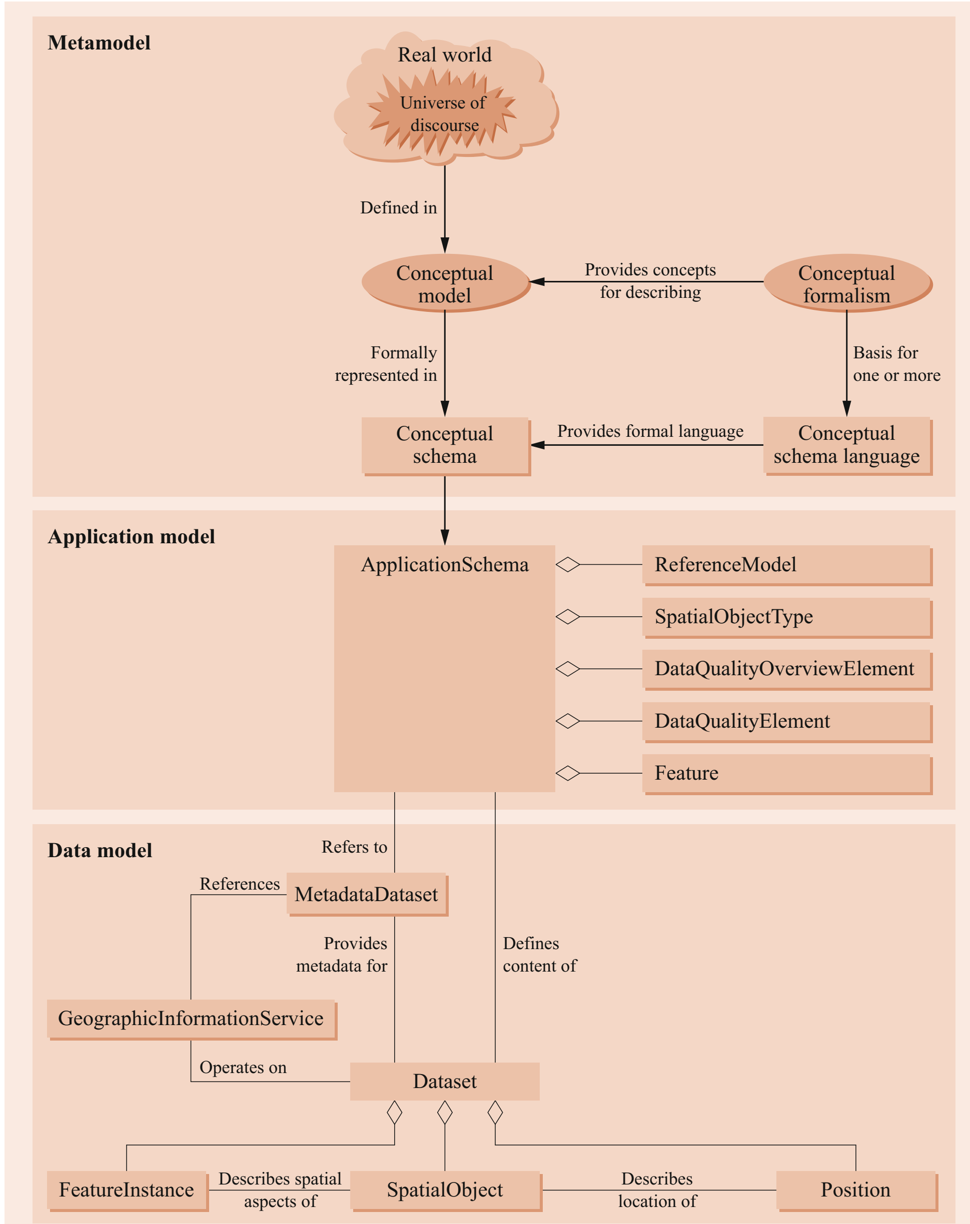
Based on the conceptual schema modeling facilities, ISO 19101-1 sets forth the process for conceptual modeling, the domain reference model, the architectural reference model, and profiles.

Conceptual modeling covers the metamodel. According to the former ISO/IEC 14481, a number of principles govern the use of conceptual modeling and the development of conceptual schemas. The most important principles are listed here:

1. The *100% principle* states that all (100%) of the relevant structural and behavioral rules about the universe of discourse shall be described in a conceptual schema. Thus, the conceptual schema defines the Universe of discourse.
2. The *conceptualization principle* states that a conceptual schema should contain only those structural and behavioral aspects that are relevant to the universe of discourse. All aspects of the physical external or internal data representation should be excluded.
3. The *self-description principle* states that normative constructs defined in the standards and profiles of the ISO 19100 series shall be capable of self-description.

The domain reference model covers the application model and the data model. The details of the domain reference

Fig. 15.7 Conceptual modeling and domain reference model. (Adapted from [18])



model were explained above. Fig. 15.7 summarizes the relations between the metamodel, the application model, and the data model.

Architectural Reference Model

The term "architecture" is used here in the context of service architecture. The architectural reference model lays out a concept for structuring the great number of geographic information services involved. The basis for the architectural reference model is the ISO/IEC Open systems environment reference model (OSE-RM) described in the withdrawn ISO/IEC TR 14252. This standard completely addresses the services for information technology. The ISO 19101-1 standard defines an *extended* open systems environment (EOSE) reference model that includes the specific services for geomatics. Fig. 15.13 provides a complete view of the IT services and the geographic information services.

Profiles

A profile of a standard defines a subset of the elements of the standard in order to meet specific needs and to eventually avoid heavy overload of functionality. Profiles of the ISO 19100 standards may combine subsets of different standards, including the ISO 19100 family and others. A special case of future profiles are functional standards that were developed outside the world of the ISO 19100 standards before the work of ISO/TC 211 started. These functional standards were planned to become profiles of the ISO 19100 family.

However, full replacement of domain-specific standards by the ISO 19100 family turned out to be the better solution, as in the case of the S-100 series, the IHO geospatial standard for hydrographic data.

Reference Model – Part 2: Imagery (ISO 19101-2)

ISO/TS 19101-2 *Reference model – Part 2: Imagery* defines a reference model for imagery [19]. Its development turned out to be necessary because Earth observation and remote sensing have become key topics of geographic information. Unfortunately, the original reference model, ISO 19101, provides guidelines primarily for vector-oriented GIS.

The key topic of ISO/TS 19101-2 is geographic imagery. This is defined as imagery whose data are associated with a location relative to the Earth and contrasts with the term "image", which is not addressed in this technical specification because of its numerous meanings in various user contexts.

ISO/TS 19101-2 offers a very broad view on geographic imagery, ranging from technical details such as encoding formats to the process of decision-making, when it claims that the ultimate application of imagery data is the support of decisions. From the perspective of ISO/TS 19101-2, most geographic imagery will never be directly accessed by humans in the future. Instead, semantic processing will be required, i.e., automatic detection of features and mining based on geographic concepts. One of the major technical challenges of imagery is the huge volume of data.

In the ISO 19100 family of standards, an image is regarded as a type of coverage that can be structured using a grid. As defined in ISO 19123, a grid is a network composed of two or more sets of curves in which the members of each set intersect the members of the other sets in an algorithmic way.

Imagery is built of a matrix of pixels. Using the terminology of the ISO/TC 211 standards, an image is a gridded coverage whose attribute values are a numerical representation of a physical parameter, e.g., light intensity.

It is useful to distinguish between imagery as used in the ISO/TS 19101-2 from the colloquial use of the term "image." As used in ISO/TS 19101-2, imagery is a representation of image data within a computer system. To view an image, a presentation process is required to convert the pixel values, called digital numbers (DN), to a viewable representation.

15.3.4 Data Modeling

Every standard developed by the ISO/TC 211 is a part of a large consistent data model. To guarantee consistency in construction and application of this data model the family of 19100-standards contains a number "data modeling standards". Those include the *conceptual schema language* (ISO 19103), *conformance and testing* (ISO 19105), *profiles* (ISO 19106), *rules for application schema* (ISO 19109), *encoding* (ISO 19118), and *services* (ISO 19119). Originally, ISO 19102 *Overview* was meant to provide a general introduction to the ISO 19100 family. The project was later canceled because it was difficult to continuously update it while standards evolved. The Internet and textbooks such as this may provide much better access to the ISO 19100 standards.

Conceptual Schema Language (ISO 19103)

Today, if experts meet to discuss the design of a computer system, they talk in a conceptual schema language [20]. Admittedly, they speak in English, French, or German, or any other language of the world, but a conversation like this remains informal and fuzzy until someone starts drawing a diagram, mostly in UML, the Unified Modeling Language, to express the ideas using the formal tools of classes or packages and their relationships. Therefore, it is essential that everybody working in this field is able to communicate in a conceptual schema language. ISO 19103 *Conceptual schema language* defines a UML profile for geographic information. This handbook assumes that the reader has basic knowledge of UML and focuses on the extensions of UML for geographic information.

Background

A conceptual schema language is based upon a conceptual formalism that provides the rules, constraints, inheritance mechanisms, events, functions, processes, and other elements that make up a conceptual schema language. For the ISO 19100 family of standards, the applicable conceptual formalism is object-oriented modeling as described by the Object Management Group (OMG) [21]. A conceptual schema language has to be capable of representing 100% of the semantics in a domain of discourse. The 100% requirement refers to the level of detail that is appropriate for modeling the domain in question. Traditional conceptual schemata, such as the entity–relationship model, cannot describe numerical or logical relationships between values of concept. Therefore, they are not able to meet the 100% requirement.

UML has become the strongest of several conceptual schema languages that have been developed over the last decades. The roots of UML were independent but similar to developments by three well-known American *software methodologists*: Booch, Rumbaugh, and Jacobson, who pooled their efforts and created a company, Rational Software Corp., that has become one of the leading developers of software engineering tools. Today, UML is an international standard: ISO/IEC 19501, prepared by the ISO/IEC JTC1/SC7 [22]. Today, Rational Software Corp. is a division of IBM.

EXPRESS is a conceptual schema language being used in the field of mechanical engineering and was standardized by

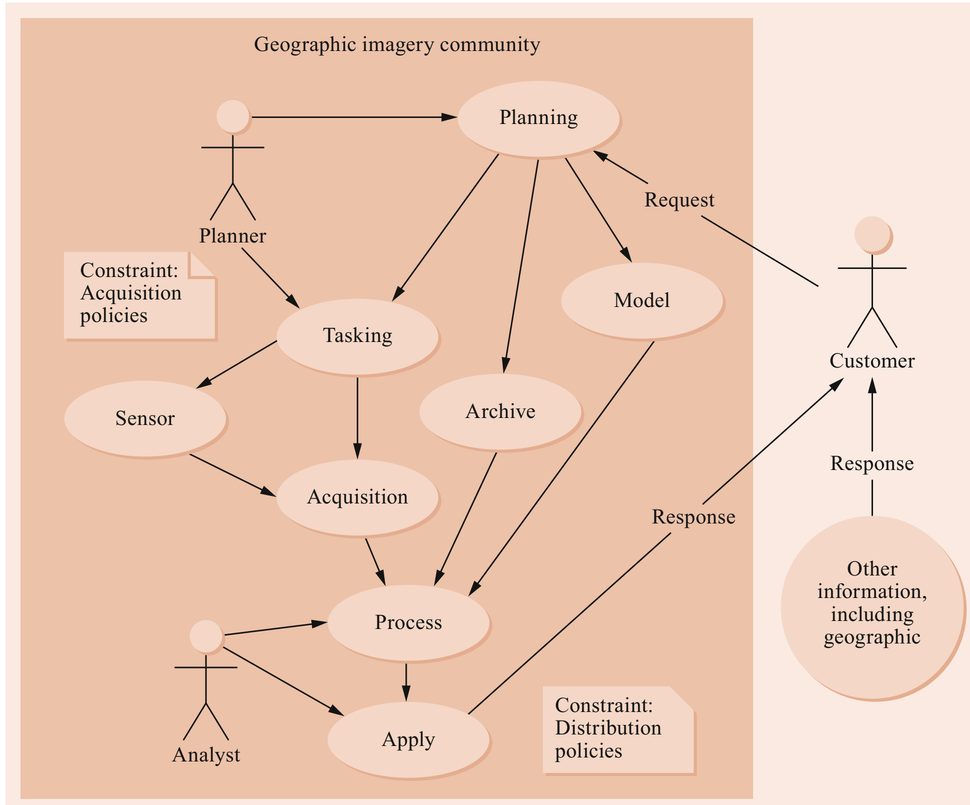


Fig. 15.8 UML use-case diagram of geographic imagery scenario. (Adapted from [19])

ISO/TC 184 *Automation systems and integration* [23]. Conceptual schemas in UML are based on graphical and lexical elements, whereas the schemas of EXPRESS primarily rely on text.

According to the standards of the ISO 19100 series, both languages are available for modeling of geographic information. UML is preferred, however, as it has turned out to be far more feasible for modeling geomatics. Therefore, this handbook uses UML as the only conceptual schema language. Today, ISO/TC 211 recommends Enterprise Architect developed by Sparx Systems as the preferred UML modeling tool.

UML Elements for Geographic Information

ISO 19103 *Conceptual schema language* requires use of UML as defined in ISO/IEC 19501. Specific rules and recommendations have been established for the following aspects: classes, attributes, data types, operations, associations, and stereotypes. In addition, naming conventions and modeling guidelines maintain the unique appearance of the whole family of ISO 19100 standards.

Classes

Normative models use class diagrams and package diagrams. Other UML diagram types, such as use-case diagrams, may be used for information (Fig. 15.8). All normative models contain complete definitions of attributes, associations, operations, and appropriate data-type definitions.

According to the ISO 19100 family, a class is viewed as a specification and not as an implementation. Attributes are considered to be abstract and do not have to be directly implemented. For each class defined according to the ISO 19100 family, its set of defined attributes together with the sets of attributes of other classes (that are accessible either directly or indirectly via associations) shall be sufficient to fully support the implementation of each operation defined for that particular class.

Attributes

All attributes must be typed, and the type must exist among the set of legal base types. A type must always be specified; there is no default type.

Relationships and Associations

A relationship in UML is a ratified semantic connection among model elements. Generalization, dependency, and refinement are class-to-class relationships. In the ISO 19100 family of standards, they are used according to the standard UML notation and usage.

Association, aggregation, and composition are object-to-object relationships. An association is used to describe a relationship between two or more classes. An aggregation is a relationship between two classes, in which one of the classes plays the role of a container, and the other plays the role of a content. A composition is a strong ag-

Generalization	A relationship between an element and the subelements that may be substituted for it
Dependency	The use of one element by another
Refinement	A shift in levels of abstraction
Association	A semantic connection between two instances
Aggregation	A part-of relationship
Composition	Strong aggregation, children are deleted if parent is deleted

Fig. 15.9 Kinds of relationships in UML. (Adapted from [20])

gregation. In a composition, if a container object is deleted, then all of its content objects are deleted as well (Fig. 15.9) (Chap. 1).

Stereotypes

Stereotypes are a method of classifying UML classes in order to augment the readability of larger UML class and package diagrams. Stereotypes indicate the context in which a class shall be applied. ISO 19103 defines 13 stereotypes as being relevant for geographic information (Table 15.3).

Conformance and Testing (ISO 19105)

If it is claimed that a data file is written in a standardized format such as TIFF, then it would be easy to test conformance with the standard. If a correct TIFF reader is able to open and display the file, then, as far as the content is concerned, both the file and the reader conform to the standard. In larger systems, however, it becomes more difficult to decide if they conform to certain standards. In order to execute a test that is independent of the system manufacturer and the user, independent institutions such as testing laboratories and accreditation bodies often become involved. Testing of conformance represents a major step during the introduction of the system to the market. ISO 19105 *Conformance and testing* sets the rules for the conformance tests of all ISO standards for geographic information [25].

Similarly to many other basics of the ISO standards for geographic information, the original rules for conformance testing were developed by the information technology community [26–28]. These rules were enhanced for use within the ISO 19100 series. Conformance testing means testing of a candidate product for the existence of specific characteristics. The testing addresses the capabilities of an implementation compared with the

1. conformance requirements as defined in each standard document and the
2. product description of the manufacturer.

The first is a test according to conformance class A, and the second is a test according to conformance class B.

ISO 19105 defines two types of conformance tests. The *basic test* provides limited testing of an implementation under test (IUT) to establish whether or not it is appropriate to perform more thorough testing. The *capability test* should exercise an implementation as thoroughly as is practical over the full range of conformance requirements specified in the *standard*.

Conformance testing does not include testing the robustness of an implementation, acceptance at the client site, or system performance.

Table 15.3 ISO 19103 (stereotypes)

<<applicationSchema>> is a package representing an application schema as defined in ISO 19109
<<baseSchema>> is a package defining items of geographic information
<<codeList>> is a flexible enumeration that uses string values for expressing a list of potential values
<<dataType>> is a set of properties that lack identity (independent existence and the possibility of side-effects). A data type is a class with no operations whose primary purpose is to hold information
<<enumeration>> is a fixed list of valid identifiers of named literal values. Attributes of an enumerated type may only take values from this list
<<featureType>> is a feature type as defined by the GFM in ISO 19109
<<interface>> is an abstract classifier with operations, attributes, and associations, which can only inherit from or be inherited by other interfaces. Other classes may realize an interface by implementing its operations and supporting its attributes and associations (at least through derivation)
<<leaf>> is a package that contains definitions, without any subpackages
<<metaclass>> A class whose instances are classes. Metaclasses are typically used in the construction of metamodels, and a metaclass exists at a higher level of abstraction; for example, FeatureType and AttributeType are metaclasses for Feature and Attribute
<<metamodel>> is a package that defines a language for expressing a model
<<union>> is a type consisting of one and only one of several alternatives (listed as member attributes). This is similar to a discriminated union in many programming languages
<<voidable>> identifies an attribute or association role as optional, i.e., a value of "void" is a valid value of the property
<<type>> is a set of abstract attributes and associations. Abstract means that their specification does not imply that they have to be concretely implemented as instance variables

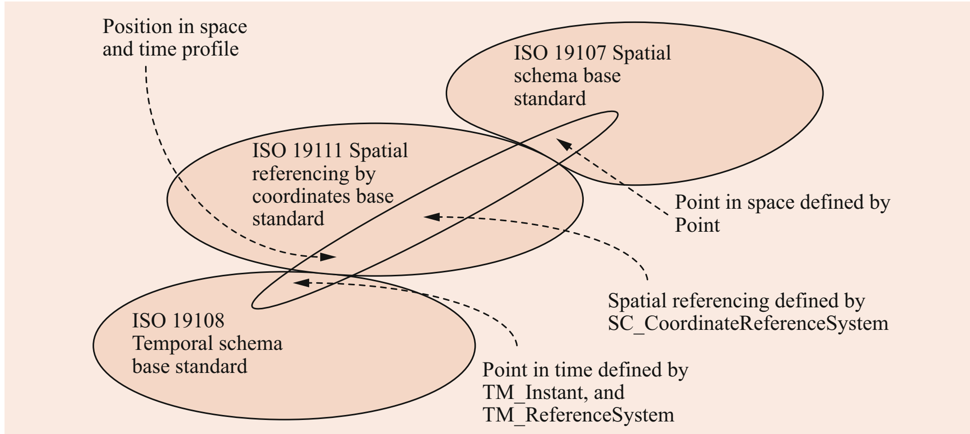


Fig. 15.10 Example of a profile using concepts and structures from more than one standard. (Adapted from [24])

All testable ISO geographic information standards contain a conformance clause that specifies all the requirements that must be satisfied in order to claim conformance to that standard. The conformance clause serves as an entry point for conformance testing. The conformance clause is hierarchically structured into the upper level of an abstract test suite, the medium level of abstract test modules, and the lower level of abstract test cases. The precise definition of the test purpose is the key statement of every test module and every test case. An example of a test purpose may be: *test the generation of a polygonal line as a sequence without self-intersection.* This is a small example, but it indicates that the abstract test suite of the major standards may become rather voluminous documents. It is for the standard's developer to create a conformance clause that includes all test methods and test cases necessary to guarantee complete conformance to the standard.

An implementation that is to undergo a conformance test is called an *implementation under test* (IUT). ISO 19105 *Conformance and testing* structures the test into four steps: preparation for testing, test campaign, analysis of results, and conformance test report. The conformance assessment process is carried out by an independent testing laboratory.

The formalized approach to testing of implementations has some important intrinsic properties. A test must be repeatable in that two or more tests of the same implementation are comparable and have mainly the same results. The test is also auditable in that the work of the independent testing laboratory may be subject to audit.

Creation of Profiles (ISO 19106)

The family of international standards for geographic information covers an immense range of possible applications with reference to the Earth. However, the complete ISO 19100 family of regulations would overload most real applications. Therefore, only a subset of the ISO 19100 standards is normally required. A profile according to ISO 19106 *Profiles* is a subset of the ISO 19100 standards (Fig. 15.10) [24].

The ISO 19100 standards provide two approaches for the creation of a specific application, profiles, and application schemas. While a profile narrows the functionality, an application schema extends it beyond the scope of a given standard in order to meet specific needs. Accordingly, an application schema is developed in two steps:

1. Definition of a profile of the ISO 19100 standards.
2. Creation of an application according to ISO 19101-1 *Reference model – Part 1: Fundamentals* and ISO 19109 *Rules for application schema* for all additional components.

A profile may become a standard on its own and is then called an international standardized profile (ISP). In the mid 1990s, the joint technical committee of the ISO and IEC (ISO/IEC JTC1) created the international standardized profile as a new type of document. To receive the status of an ISO standard, the document describing a profile must follow the procedures for the development of an international standard. As a result, it receives its own ISO number. Profiles defined according to the ISO Profile standard may become this type of international standard.

One may argue that, in times of fast computers and the availability of enormous amounts of disc space, it is not necessary to artificially narrow the options offered by a large environment such as the ISO 19100 family of standards. However, the idea of the introduction of profiles is to promote better interoperability between systems by restricting the choices. It is easier for users and system suppliers to agree on a smaller set of common standards than on a large set.

How is a profile defined?

- A profile is a subset of the base standards of the ISO 19100 family of standards or other information technology standards.
- A profile determines how they are used together.
- A profile explains the usage details as far as required.

ISO 19106 distinguishes between two types of profiles:

- Conformance level 1 designates a profile that is purely derived from elements of the ISO 19100 family of standards and possibly other ISO standards.
- Conformance level 2 specifies profiles that integrate elements of non-ISO standards with elements from the ISO standards.

The development of conformance level 2 profiles is permitted because the existing suite of standards cannot yet claim to meet *all* requirements. This is despite the great effort that has been put into a generalized and comprehensive approach towards geographic information. A profile of conformance level 2 cannot become an ISO standard of the international standardized profile type and will not be generally interoperable.

Rules for Application Schema (ISO 19109)

The ISO 19100 standards address the full range of geographic information. A certain application only uses a subset of the available standards, combined with many additional details that are beyond the scope of the abstract standards. An example is a feature catalogue for a given application.

The ISO 19100 standards provide two approaches to creating a specific application: profiles and application schemas.

The kernel of ISO 19109 is the definition of a geographic feature. A feature stands for anything in the real world; for example, a feature can be a single corner stone of a land parcel, a whole country, a digital elevation model, or a satellite image. In order to integrate a feature into the models of geographic information in a homogeneous way, ISO 19109 defines the GFM [16].

The GFM defines an abstract feature with attributes and operations. Attributes contain all the static information, such as the quality of the feature or its geometric properties (point, curve, surface, or solid). Operations contain information about the change of a feature according to external influences, such as a road being closed in winter, or a road being displayed only within the scale range 1:5000 to 1:25 000. This change is also called the behavior of the feature.

As these examples show, features can differ in importance and size. In practice, this often leads to a hierarchical grouping of features; for example, public buildings and private buildings are both *buildings*. The GFM allows for the construction of a generalization tree where the feature types (class GF_FeatureType) public and private buildings are specializations within the feature type building.

Generally, features reside independently in the dataset. In the case of a three-dimensional dataset, the relations between features can be computed in all spatial dimensions. The over, under, or level situation at intersections is a particularly important case for mapping and network computation. A three-dimensional dataset implicitly contains the necessary information. A two-dimensional dataset does not allow for these computations. To supply the missing information, the GFM includes associations between features that contain the information regarding which feature is above the other, or if both are at the same level. In this case, the association type (class GF_AssociationType) would be *intersection*, and the association roles (class GF_AssociationRole) would be under, over, or level, respectively.

In some cases, it may be advisable to impose constraints on the definition of a feature. A theoretical example may be that a feature of type curve must not have more than 1000 points. The GFM allows the formulation of these constraints (class GF_Constraint).

An application schema is usually created by the definition of the features. All details are collected in a feature catalogue. The methodology for building feature catalogues is covered by ISO 19110.

Encoding (ISO 19118)

The concept of the exchange of geographic information datasets is standardized in ISO 19118 *Encoding*. It defines a system-independent data structure for transport and storage and normatively demands the usage of XML for encoding. Presently, the ISO 19100 family of standards defines two XML encoding rules: ISO 19136 and ISO/TS 19139 *Metadata – XML schema implementation* for the encoding of metadata. Either rule set may be applied depending on the use case.

The primary goal of the ISO 19100 family of standards is to enable full interoperability between heterogeneous geographic information systems. To achieve this goal, two fundamental issues need to be resolved. The first issue is to define the semantics of the content and the logical structures of geographic data. This is achieved by implementing the same application schema. The second issue is to define a system-dependent and platform-independent data structure that can represent data corresponding to the application schema; for example, equal semantics and logical structures can be guaranteed by using the same feature catalogue in the two systems involved. A system-independent and platform-independent data exchange format must be created according to ISO 19118 *Encoding*.

A Model for Data Exchange

The generic model for the data exchange assumes two systems that run on different computer platforms and have different application schemas. The transfer of a dataset between such systems is modeled in six steps.

Step 1: The source system translates its internal data into a data structure that is according to the common application schema.

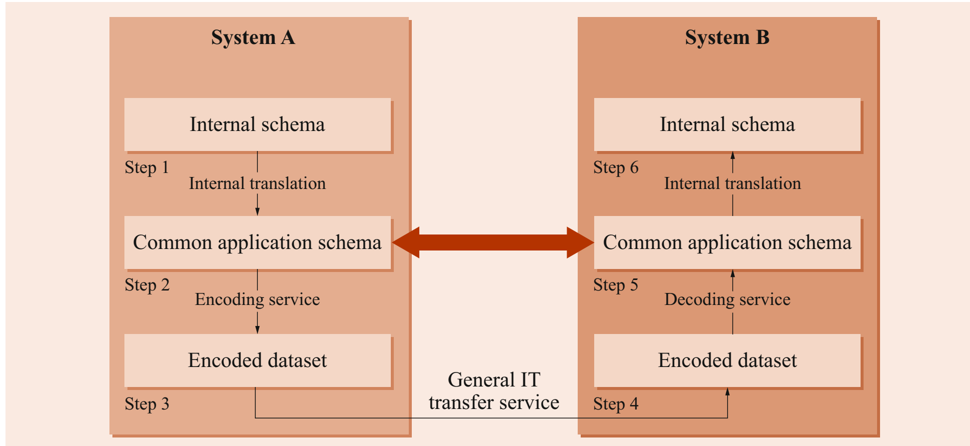


Fig. 15.11 A model for data exchange, simplified. (Adapted from [29])

Step 2: An encoding service applies the encoding rules to the data, creating a file or transferring the data to a transfer service.
Step 3: The source system invokes a transfer service to send the encoded dataset to the destination system.
Step 4: The destination system receives the dataset.
Step 5: The destination system applies the inverse encoding rule to interpret the encoded data.
Step 6: The destination system must translate the application schema specific data into its internal database.

Steps 2 and 5 are standardized in ISO 19118, which describes the encoding and the decoding of the dataset. These two steps are handled by an encoding service that is a software component that implements the encoding rule and provides an interface to encoding and decoding functionality.

The encoding rule specifies the data types to be converted, as well as the syntax structure and coding schemes used in the resulting data structure, preferably an XML document.

Steps 3 and 4 use general information technology transfer services to send and receive data.

Steps 1 and 6, the internal translations within the source and destination systems, are outside the scope of this standard (Fig. 15.11).

Encoding Using Extensible Markup Language (XML)

As the XML is recommended for data exchange, ISO 19118 *Encoding* defines a mapping for application schemas that are written in the UML to the corresponding data structures of XML. There are two views to that mapping:

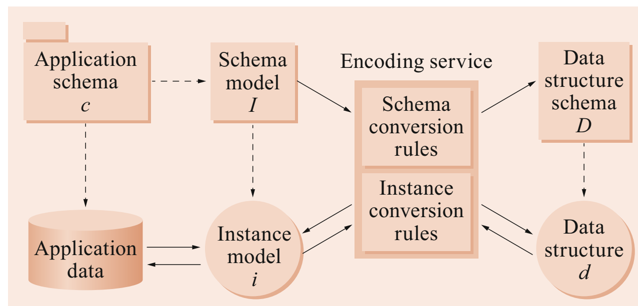


Fig. 15.12 XML-based conversion rules. (Adapted from [29])

1. The *abstract view* is independent from any dataset. This is the *application schema*. It contains the classes and other components, such as attributes and associations, that fully describe all geographic feature types belonging to a certain application and their relations. A representation of the application schema is a feature catalogue. In XML the application schema is expressed as an XML schema document (XSD), which contains the rules according to which an XML document must be built.
2. The *implementation view* addresses a specific geographic dataset. This is the *application data*. It is, for example, a map compiled according to a feature catalogue. As every geographic feature is an *instance* of one class of the application schema, the dataset is addressed as the *instance model* in ISO 19118. For the purpose of exchanging the dataset, the instance model is encoded in an XML document. The structure and the tags of that XML document are built according to the XML XSD (Fig. 15.12).

Services (ISO 19119)

The IT community decomposes computer software systems by using five different viewpoints. One of them, the computational viewpoint, addresses services. ISO 19119 *Services* provides the framework to structure services in the context of geographic information. In the case of geomatics, services may be pure information technology services, such as querying a database, or specific geographic information services, such as finding a location in a coordinate reference system (CRS).

Services are defined as the capability to provide for manipulating, transforming, managing, or presenting information. A special case is the service interfaces that are the boundaries across which services are invoked and across

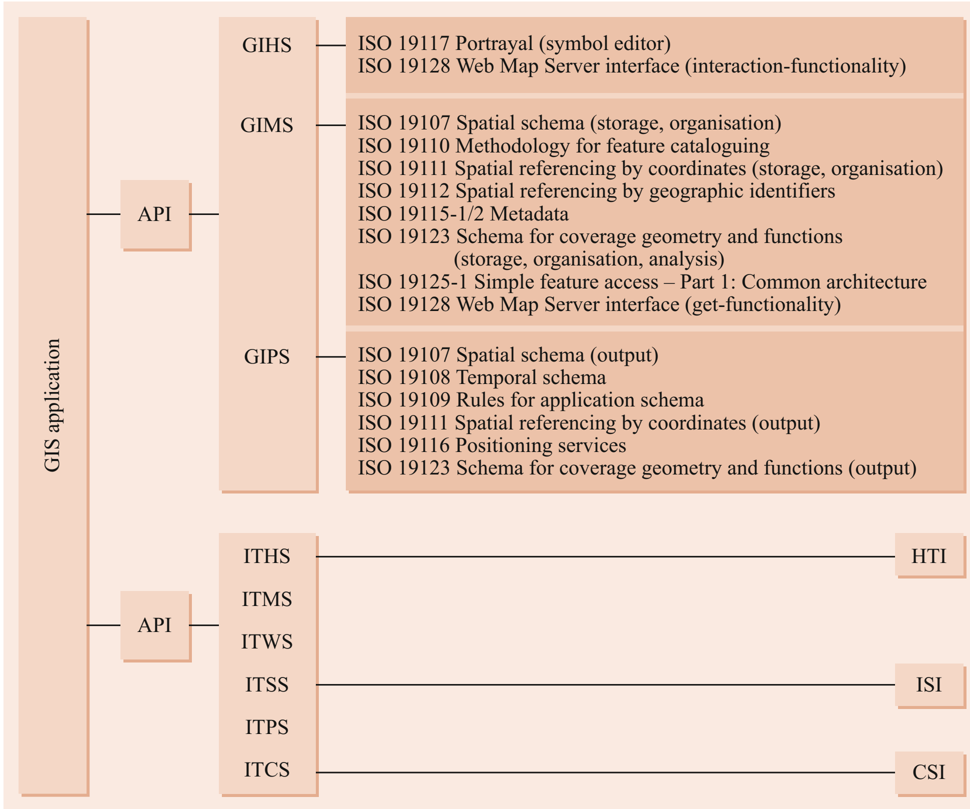


Fig. 15.13 Architectural reference model showing the services and their relations; mapping of the ISO 19100 standards to the extended OSE categories (Adapted from [18, 30])

which data are passed between a service and an application, external storage device, communications network, or a human being. Following these definitions, the ISO 19100 standards fall into two categories: infrastructure standards and service standards. Infrastructure standards are guidelines that are applicable for all other standards. Examples are ISO 19101-1 *Reference model – Part 1: Fundamentals* and ISO 19104 *Terminology*. With few exceptions, all standards with numbers between ISO 19107 and ISO 19119 are service standards.

ISO 19101-1 *Reference model – Part 1: Fundamentals* sets forth the conceptual modeling, the domain reference model, and the architectural reference model. The latter embraces the IT services as defined in the Reference model for Open distributed processing (RM-ODP) and standardizes the extended open system environment reference model (EOSE-RM) for the specific services for geographic information. The general open system environment reference model (OSE-RM) is standardized in ISO/IEC 10746 [31].

Fig. 15.13 summarizes all IT and geographic information services. Without claiming completeness, the diagram also relates the ISO 19100 standards to the three service categories of geographic information. The acronyms are explained below.

The OSE-RM defines six types of services and five types of interfaces through which the services communicate with other applications or peripheral devices:

- IT service types
  1. Human interaction service (ITHS). An example is the interaction of an operator with the graphic interface while logging on.
  2. Model management service (ITMS). An example is the software-supported creation of an UML diagram while designing a geographic application.
  3. Workflow/task service (ITWS). An example is the setting of a later start time for an overnight run of a large processing job.
  4. System management service (ITSS). An example is the creation of a new user with access rights to a geographic database.
  5. Processing service (ITPS). An example is the calculation of the average density of population in a given dataset.
  6. Communication service (ITCS). An example is the provision of map data over the Internet.
- Interface types
  1. Application programming interface (API). All services communicate via an API with GIS applications.

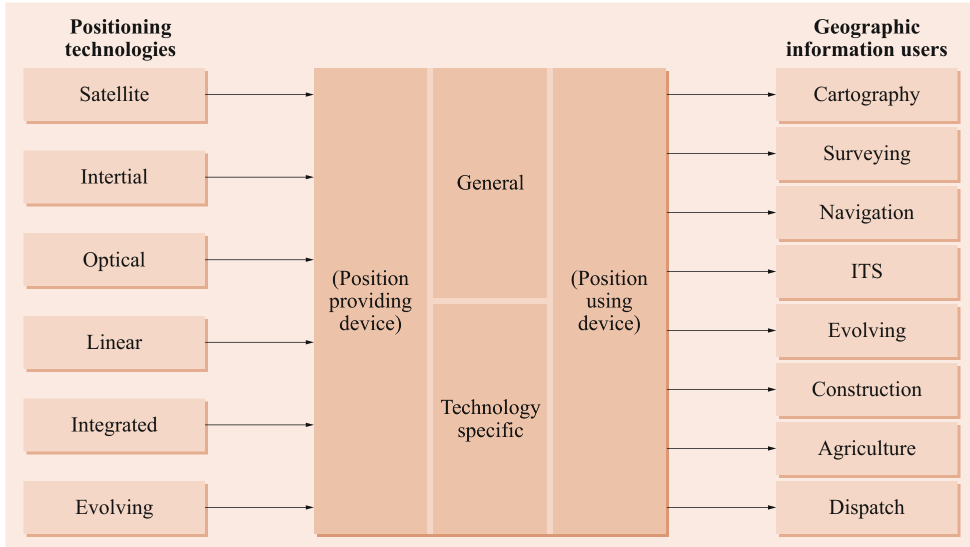


Fig. 15.14 Overview of application cases of the positioning services (Adapted from [32]) (ITS – intelligent transport systems)

2. Human technology interface (HTI). The best-known examples are the window systems on the screen.
3. Information service interface (ISI). The ISI establishes the link to all data sources, such as the databases from where data shall be retrieved.
4. Communications service interface (CSI). The CSI is the link between the computer and the network.
5. Network-to-network interface. This is the network itself.

In addition, the EOSE-RM includes the following services. Theoretically, every IT service has an equivalent geographic information service. In fact, only three categories are relevant for geographic information:

- GI service types
  1. Geographic information human interaction service (GIHS). An example is the interaction of an operator with the graphic interface while picking a line.
  2. Geographic information model management service (GIMS). An example is the storage of a geographic feature according to an application schema.
  3. Geographic information processing service (GIPS). An example is the output of a map to a screen.

Positioning Services (ISO 19116)

Positioning services are computer techniques that deliver positions of an object relative to the Earth or to some other position. These techniques include Global Navigation Satellite Systems (GNSS), inertial systems, and total stations. ISO 19116 *Positioning services* standardizes a data model for the basic information independent of the system type, and a group of operations to handle those data. A system-specific section is dedicated to GNSS only.

Finding a position is no longer the domain of a skilled navigator or experienced surveyor, but any user can still suffer from the great variety of different and incompatible interfaces among positioning systems. ISO 19116 standardizes this interface and isolates the client from the multiplicity of protocols.

Within the ISO 19100 family, positioning services are among the processing services identified in ISO 19119.

Fig. 15.14 gives an overview of the application cases of the positioning services. The diagram covers the standardized types, including some future conceptual types.

15.3.5 Geometry

Traditionally, graphic data fall into the two categories of vector and raster. This approach is reflected in the way the ISO 19100 standards are partitioned. The ISO 19100 standards use the more general term *gridded data* instead of raster.

ISO 19107 *Spatial schema* depicts a fairly complete world of vector data. ISO 19125-1 *Simple features access – Part 1: Common architecture* is a subset of ISO 19107 using different terminology because of historic reasons.

ISO 19123 *Schema for coverage geometry and functions* describes coverage data in a general sense. Coverages are often considered to be a way of integrating the worlds of vector and raster data, thus overcoming the dichotomy originally imposed by hardware restrictions. However, coverages are restricted to three spatial and one temporal dimension. Later, they shall handle *n*-dimensional space. ISO/TS 19130-1 *Imagery sensor models for geopositioning – Part 1: Fundamentals* describes the models needed to relate remotely sensed imagery data to the Earth. ISO/TS 19129 *Imagery, gridded and coverage data framework* is a framework stan-

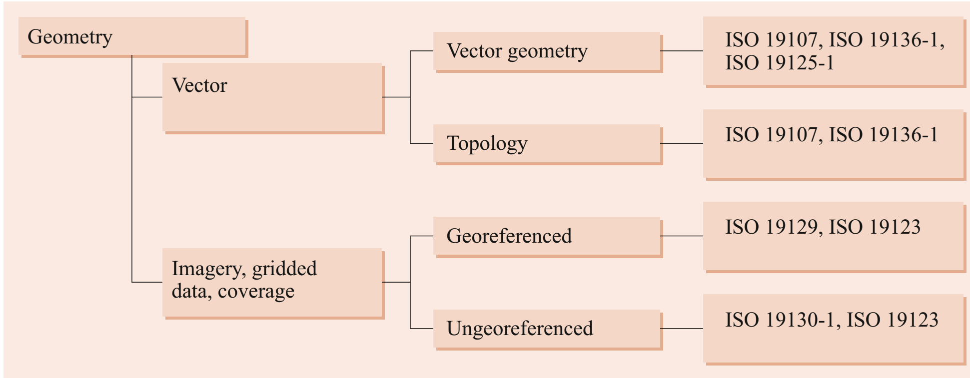


Fig. 15.15 Relation of the vector data and imagery standards of ISO 19100

dard that describes the concepts of gridded data, lists all required elements, and fills some gaps left by the other standards.

This section will start with discussing the reference systems followed by vector and raster and concludes with the topic of visualization.

Relation Between Geometry Standards

The 19100 family of standards comprise six geometry and imagery standards. The diagram above (Fig. 15.15) shows the relations between them.

Although a less abstract way of describing the structures may be desirable in some cases, the terminology of the conceptual schema language UML will be used in order to be consistent with the UML models of the ISO 19100 standards. In most cases, the models consist of a hierarchy of classes. A high-level class describes a graphical element in a very general sense; for example, a Geometry can be any geometry object of the dataset, such as a point, a curve, a surface, or a solid. A low-level class describes a specific geometry, such as an arc or a spline. The whole model contains an abstract view of the real world, and its only purpose is to address the classes that could possibly be present in a dataset. If a real dataset is created, it uses the classes supplied by the standard. The elements that form the real dataset are the instances of the abstract classes; for example, the instances of the class Surface, may be the property parcels $1, 2, \ldots, n$ in the dataset.

Position

ISO 19107, ISO 19123, and possibly others in the future, have an integrated concept for dealing with positions in n-dimensional space (Fig. 15.16). ISO 19125-1 supports only two dimensions and uses a different method.

The concept based on ISO 19107 will be described first. An n-dimensional set of coordinates is called a DirectPosition, such as (x, y, z). Obviously, a point object (class Point) has one set of coordinates and, thus, one DirectPosition. Curves, surfaces, and solids have more than one set of coordinates and, consequently, more than one DirectPosition.

The name DirectPosition is derived from the *direct* identification of the position. Alternatively, a position can be defined by the position of another *point object* (class Point). This is called an *indirect* identification of a position. An example would be a parcel boundary, which is mostly a line string. The positions of the line string are the corner stones of the parcel, which are point objects themselves. Thus, the coordinates of the line string are the coordinates of the point objects.

To better handle numerous DirectPositions in a dataset, the positions of a curve, a surface, or a solid are collected in an ordered manner, which is called a point array (class DirectPosition [ ]). To avoid redundancy and unnecessary use of storage volumes, the point array only stores pointers and no coordinates. The pointers may point to DirectPositions or point objects. The pointers within a point array (class DirectPosition [ ]) are alternatively called controlPoints. The class storing the flag with direct or indirect reference is called DirectPosition.

The ISO 19123 concept of a coverage relates a set of attributes to a position within a bounded space. From this perspective, the basic element of a coverage is a pair of data collections; one data collection denotes a position such as (x, y), and the other data collection denotes the attributes at this position such as temperature and soil type. ISO 19123 refers to ISO 19107 for the details of positions.

This pair of data collections at one point is called a point–value pair (class CV_PointValuePair). According to ISO 19107, the position of a point–value pair is always a point object. In other words, the DirectPosition of ISO 19107 contains the coordinates of a position, such as (x, y), and the point object (class Point) uses this DirectPosition. The point–value pair of ISO 19123 (class CV_PointValuePair) uses the point object of ISO 19107 for position information.

A common pattern of points is a quadrilateral grid. Examples of this would be a digital image and a simple type of digital elevation matrix. For reasons of storage and computational efficiency, the quadrilateral pattern of points has some special terminology and structure. The grid addressed as a whole is called a *grid values matrix* (class CV_GridValuesMatrix).

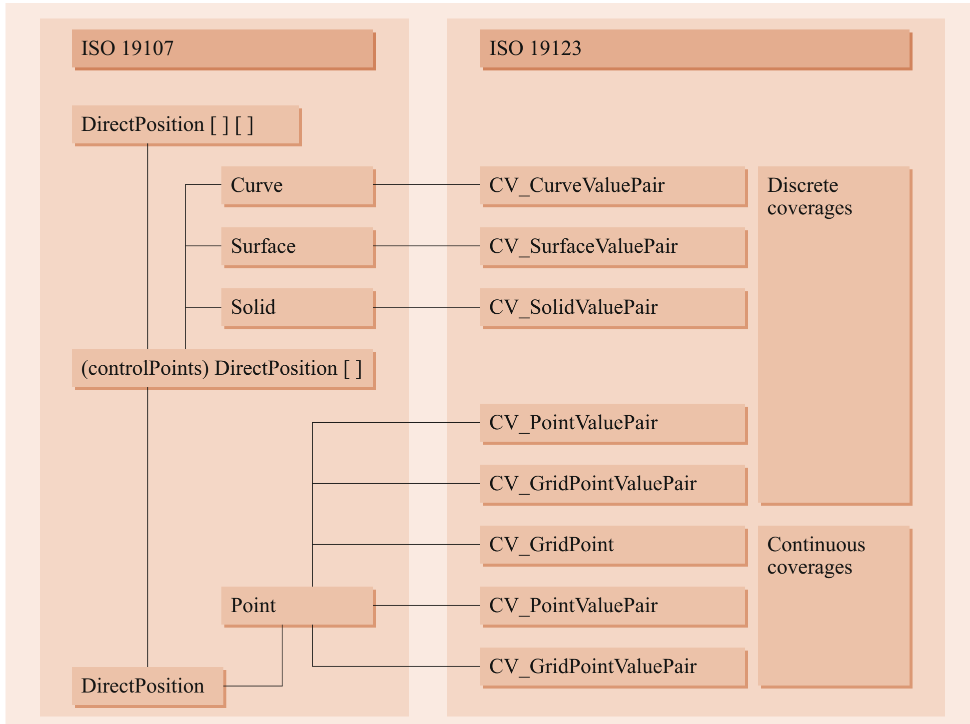


Fig. 15.16 Relations of position descriptions

The intersections of the grid lines are called *grid points* (class CV_GridPoint). According to ISO 19107, the position of each grid point is always a *point object*.

ISO 19107 states that a grid can also be defined as a quadrilateral pattern of simple point positions or point objects (class Point), as in ISO 19123. The grid is called a *point grid* (class DirectPosition [ ]) and consists of a number of parallel-placed point arrays (class DirectPosition [ ]).

A knot is a special type of point that is only used for the description of some types of splines; for example, a Bézier spline is defined by four or more points where the curve passes through the first and last point and is only controlled in shape by the points in between. Those points are the knots of the Bézier spline.

Each dimension requires its own sequence of knots. Thus, a spline curve has one sequence, and a spline surface has two sequences.

Reference

This section addresses spatial referencing by coordinates and parametric values (ISO 19111, ISO 19111-2), geodetic codes and parameters (ISO 19127), geodetic references (ISO 19161-1), well-known text for coordinate reference systems (ISO 19162), geographic point locations (ISO 6709), spatial identifiers (ISO 19112), and place identifier architecture (ISO 19155).

Coordinate Reference (ISO 19111, 19111-2)

ISO 19111 *Referencing by coordinates* models coordinate systems and coordinate operations for the ISO geomatics standards. This model contains the horizontal and vertical coordinate references for geographic features. According to this standard, coordinate reference systems may be one, two, or three-dimensional and do not change over time.

The third dimension may be a nonspatial parameter such as pressure. However, this case is addressed in a separated standard, namely ISO 19111-2 *Spatial referencing by coordinates – Part 2: Extension for parametric values* [33, 34].

Both standards were jointly developed by ISO/TC 211 and the Open Geospatial Consortium (OGC).

Coordinate Reference System

A coordinate reference system (CRS) is a coordinate system that has a reference to the Earth by a so-called datum. ISO 19111 standardizes the details in order to fully define a coordinate reference system.

A CRS may be geodetic, vertical, engineering, or image.

A geodetic CRS using 2-D ellipsoidal coordinates (latitude and longitude) is used when positions of features are described on the surface of the ellipsoid. Vertical CRSs make use of the direction of gravity to define the concept of height or depth.

In addition, three more subtypes of CRS shall be distinguished: derived, projected, and compound.

A derived coordinate reference system is one which is defined by applying a coordinate conversion to another coordinate reference system. A projected coordinate reference system is derived from a base geodetic CRS by applying a coordinate conversion known as a map projection to latitude and longitude ellipsoidal coordinate values (Fig. 15.18).

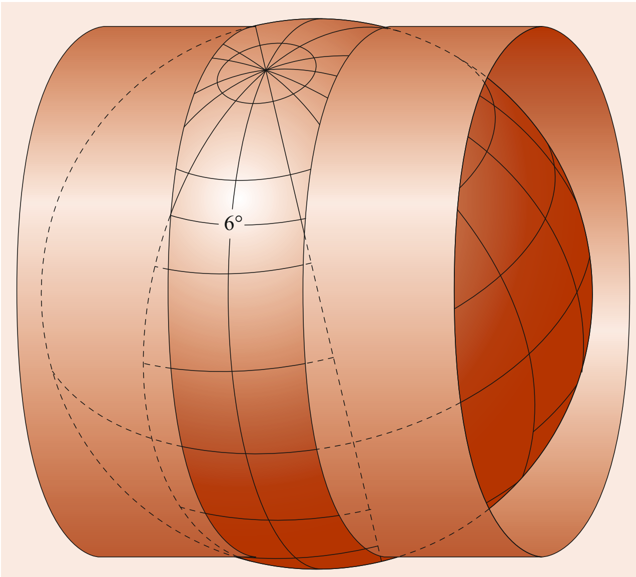


Fig. 15.17 Ellipsoidal and projected coordinate reference system (CRS). Example: Universal Transverse Mercator (UTM) projection with strips 6° wide in longitude. (Adapted from [35])

A reference that is defined by two different coordinate reference systems, one of which is an elevation system, is called a *compound* CRS (Fig. 15.19). A compound CRS is usually used to provide an independent reference for horizontal and vertical coordinates.

For spatial coordinates, a number of constraints exist for the construction of compound CRSs. Valid combinations are:

- Geodetic 2-D + vertical
- Geodetic 2-D + engineering 1-D (near vertical)
- Projected + vertical
- Projected + engineering 1-D (near vertical)
- Engineering (horizontal 2-D) + vertical
- Engineering (1-D linear) + vertical.

A CRS is called fully defined if it has a datum and a coordinate system. Historically, the datum was often the position of an observatory near the center of a country where the position and the orientation of the coordinate system are defined in relation to the physical reality of the Earth. The prime meridian defines the origin of the coordinate values of the first ellipsoidal axis.

If it is an ellipsoidal system, the three coordinates are longitude, latitude, and ellipsoidal height. In most applications, longitude and latitude are referred to as geographic coordinates. Because ellipsoidal heights are sometimes difficult to relate to the topography of the Earth's surface, the third coordinate is usually not related to the ellipsoid. Instead, it could be given in an independent elevation reference system. This method is acceptable, because in small-scale and medium-scale applications, the decrease of the height accuracy caused by introducing a second reference system is insignificant compared with the accuracy of the height values.

Datum and Coordinate Systems

ISO 19111 defines eight types of coordinate systems (Table 15.4).

A geodetic datum gives the relationship of a coordinate system to the Earth. It is geodetic, vertical, engineering, or image (see above). In many cases, it requires an ellipsoid definition. A prime meridian defines the origin from which longitude values are specified. Most geodetic datums use Greenwich as their prime meridian.

Engineering datums are used in a local sense, often being applied to platforms such as satellites and ships. An image CRS is an engineering CRS applied to images with small extension, such as an affine transformation.

Coordinate Operations

The details of the specification of a CRS become important if datasets with a reference to two or more CRSs are processed in the same environment. The coordinates of all features must be made available in the same CRS. Consequently, positions given in other systems must be converted or transformed. The standard describes the information required to change coordinate values from one CRS to another by operations.

The operation is a generalized term for conversion and transformation (supertype). The standard also sets the frame for changing coordinates between two different CRSs and distinguishes between *conversion*, where the involved CRSs have the same datum, and *transformation*, where the involved CRSs have different datums. In addition, the standard defines the concatenated operation and the pass-through operation.

A *conversion* changes coordinates from one coordinate system to another based on the same datum. In a coordinate conversion, the parameter values are exact. The coordinate conversion includes the map projection, the coordinate conversion of ellipsoidal coordinates to three-dimensional Cartesian coordinates, unit changes, and the shifting of the origin towards a local grid (Fig. 15.17). A map projection converts three-dimensional ellipsoidal coordinates (excluding height) to two-dimensional Cartesian coordinates. In all cases, the conversion is based on exact formulas that are well known in the scientific literature and that are described in Chap. 8. An example for a map projection is the conversion between ellipsoidal coordinates on the Hayford ellipsoid to projected coordinates in the Universal Transverse Mercator (UTM) (Hayford ellipsoid), and vice versa. Elevations are not taken into account.

A *transformation* changes coordinates between different CRSs with different datums. The shift, rotation, and scaling between different CRSs are derived from identical nonerror-free defined points. The orientation of global CRSs is always

15

Table 15.4 Coordinate system subtypes. (Adapted from [33])

CS subtype	Description	Used with CRS type(s)
Affine	Two or three-dimensional coordinate system with straight axes that are not necessarily orthogonal	Engineering image
Cartesian	Two or three-dimensional coordinate system that gives the position of points relative to orthogonal straight axes. All axes shall have the same unit of measure	Geodetic projected engineering image
Cylindrical	Three-dimensional coordinate system consisting of a polar coordinate system extended by a straight coordinate axis perpendicular to the plane spanned by the polar coordinate system	Engineering
Ellipsoidal	Two or three-dimensional coordinate system in which position is specified by geodetic latitude, geodetic longitude, and (in the three-dimensional case) ellipsoidal height	Geodetic
Linear	One-dimensional coordinate system that consists of the points that lie on the single axis described. Example: usage of the line feature representing a pipeline to describe points on or along that pipeline This international standard only lends itself to be used for simple (= continuous) linear systems. For a more extensive treatment of the subject, particularly as applied to the transportation industry, refer to ISO 19133	Engineering
Polar	Two-dimensional coordinate system in which position is specified by distance from the origin and the angle between the line from the origin to the point and a reference direction	Engineering
Spherical	Three-dimensional coordinate system with one distance, measured from the origin, and two angular coordinates. Not to be confused with an ellipsoidal coordinate system based on an ellipsoid *degenerated* into a sphere	Geodetic engineering
Vertical	One-dimensional coordinate system used to record the heights (or depths) of points dependent on the Earth's gravity field. An exact definition is deliberately not provided, as the complexities of the subject fall outside the scope of this specification	Vertical

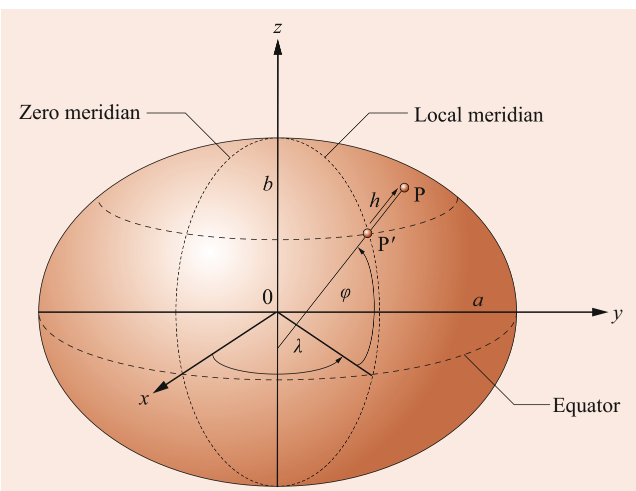


Fig. 15.18 Ellipsoidal and Cartesian coordinate system (a – semimajor axis, b – semiminor axis, λ – longitude, φ – latitude, h – ellipsoidal height of point P) (Adapted from [33])

an estimation, because it is based on a number of well-defined (but differently selected) reference points, where the residuals are minimized according to an adjustment method. Because the exact geometric relation between the CRSs is not exactly known, a transformation can only be performed with an accuracy level that corresponds to the lowest accuracy of the definition of the system parts itself; for example, if coordinates given in WGS84 (WGS84 ellipsoid) are needed in the European ETRS89 system (GRS80 ellipsoid), they must be transformed.

The concatenated operation puts two or more operations in a sequence, for instance, the three-step operation conversion–transformation–conversion. An example is the change of point coordinates from the German datum to ETRS89, which must be performed by converting the coordinates to the Bessel ellipsoid, followed by transforming them to the GRS80 ellipsoid, and finally converting them to ETRS89 projected coordinates.

Coordinate Reference: Parametric values (ISO 19111-2)

ISO 19111-2 *Spatial referencing by coordinates – Part 2: Extension for parametric values* defines the mechanism to combine a horizontal spatial coordinate reference system with a nonspatial third dimension. The *pressure altitude* that is universally used in aviation is a coordinate of this kind. This is distinct from the situation where a measure of the parameter is present as an attribute of the spatial object, such as the air temperature at a given position.

The spatial and the nonspatial CRS form a compound CRS, as defined in ISO 19111. This compound CRS may also be combined with one or more temporal CRSs according to ISO 19108.

Geodetic Codes (ISO 19127)

Within the ISO 19100 family of standards, CRSs are addressed in ISO 19111 *Referencing by coordinates*, but they do not standardize a specific CRS. Therefore, the ISO technical specification ISO 19127 *Geodetic codes and parameters* bridges the gap between the abstract frame of ISO 19111 and practical needs.

Within the geographic information community, many references exist that definie geodetic codes and parameters, none of which are in full compliance with ISO 19111. ISO 19127 *Geodetic codes and parameters* provides the required guidance to apply ISO 19111 *Referencing by coordinates* in an appropriate manner.

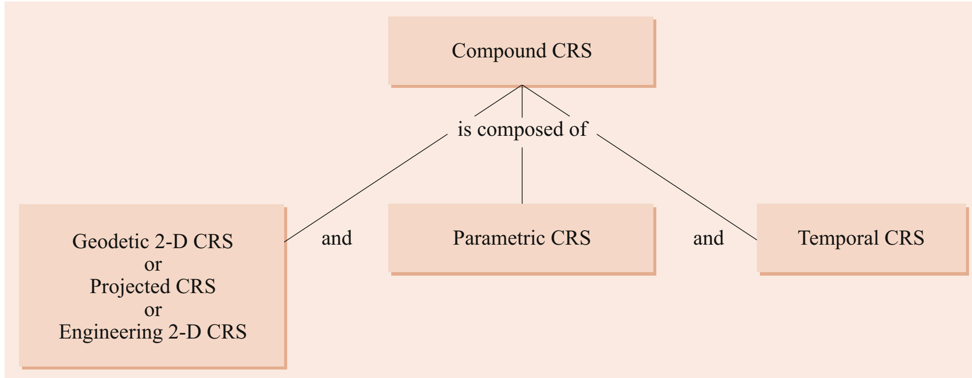


Fig. 15.19 Compound coordinate reference system including spatial (ISO 19111), parametric (ISO 19111-2), and temporal (ISO 19108) CRSs. (Adapted from [34])

The mechanism in the ISO 19100 family for creating publicly available lists of codes and parameters is a registry (ISO 19135-1). ISO 19127 provides rules for the creation and maintenance of registers for geodetic codes and parameters [36].

Geodetic References – Part 1: ITRS (ISO 19161-1)

Satellite navigation and geo-information systems of larger areas have fostered the demand for global reference systems. This requirement has become particularly crucial since global systems of today allow accuracies of a few centimeters.

A Technical Report published as ISO 19161-1 *Geodetic references – Part 1: International terrestrial reference frame (ITRS)* confirmed the need for a standardized definition, realization, and access to global terrestrial reference systems (GTRS).

The ISO 19161-1 *Geodetic references – Part 1: ITRS* recommends the International Terrestrial Reference System as the global terrestrial reference system for the whole geoscience community [37]. Other realizations of the GTRS are WGS-84 and ETRS89. The latter was adopted by the European Commission based on the INSPIRE directive within its geographical scope, i.e., within continental Europe, in contrast to the EU regions outside, such as the Caribbean EU islands.

Well-Known Text for CRS (ISO 19162)

The ISO 19162 *Well-known text representation of coordinate reference systems* defines the structure and content of a text string implementation of the abstract models described in ISO 19111:2019 *Referencing by coordinates* and ISO 19111-2:2009 *Spatial referencing by coordinates – Part 2: Extension for parametric values*. The string defines frequently needed types of coordinate reference systems and coordinate operations in a self-contained form that is easily readable by machines and by humans [38].

The following example shows a map projection using the well-known text representation.

```
CONVERSION["UTM zone 33N",
  METHOD["Transverse Mercator",
    ID["EPSG",9807]],
  PARAMETER["Latitude of natural origin",0,
    ANGLEUNIT["degree",0.0174532925199433],
    ID["EPSG",8801]],
  PARAMETER["Longitude of natural origin",+15,
    ANGLEUNIT["degree",0.0174532925199433],
    ID["EPSG",8802]],
  PARAMETER["Scale factor at natural origin",
    0.9996,
    SCALEUNIT["unity",1.0],ID["EPSG",8805]],
  PARAMETER["False easting",500000,
    LENGTHUNIT["metre",1.0],ID["EPSG",8806]],
  PARAMETER["False northing",0,
    LENGTHUNIT["metre",1.0],ID["EPSG",8807]]]
```

Point Location (ISO 6709)

The latest version of this standard, ISO 6709:2008, defines a model to describe a three-dimensional position by longitude, latitude, and height or depth in a computer-readable and a human-readable form. The previous version, ISO 6709:1983, was not fully fit for geographic information in that it did not support many requirements, such as an XML representation and the depth needed for hydrography [39].

The model set in ISO 6709:2008 allows for processing of point coordinates in a generalized manner in order to include existing standards such as:

- ISO 6709:1983 *Standard representation of latitude, longitude and altitude for geographic point locations*
- DCMI point encoding scheme (Dublin core metadata initiative)
- KML (keyhole markup language)
- GeoVRML (geographical data using the virtual-reality modeling language)
- GML Point profile (geography markup language).

The number of this standard does not comply with the ISO 19100 family, because the original ISO 6709:1983 *Standard* *representation of latitude, longitude and altitude for geographic point locations* was developed by ISO/IEC JTC1. In the year 2002, maintenance of ISO 6709 was transferred to ISO/TC 211.

Registry of Geographic Point Location (ISO 19145)

Geographic points may be encoded in many different ways. ISO 6709:2008 *Standard representation of geographic point location by coordinates* defines the model for this encoding.

The automated processing of point encodings requires adequate online access, which is technically established using a register. ISO 19145 *Registry of representations of geographic point location* defines a registry for point locations according to ISO 19135-1 *Procedures for item registration*.

ISO 19145 does not define a register of CRSs but is concerned with the manner a geographic point location according to ISO 6709 is physically represented in a record or part of it.

For the purpose of unequivocal storage and retrieval of point data in a register, ISO 19145 also defines an XML implementation of ISO 19135-1 and its specialization for point locations [40].

Geographic Identifiers (ISO 19112)

The position of geographic features is often described through their spatial relation to other geographic features. This relation may be a containment, a local measurement, or a loose relation. An example of containment is a city within a province; an example of a local measurement is the distance to the next major road intersection, and an example of a loose relation is a restaurant *between the museum and city hall*. A typical application is the partitioning of an area using postal codes. These types of positions are addressed in ISO 19112 *Spatial referencing by geographic identifiers* [41].

All positions are related to a spatial reference system. The spatial reference system comprises a subdivision of a territory such as the hierarchy province–city–address. The core element of the reference system is a gazetteer that adds a descriptive position to every geographic feature in the territory.

A gazetteer is a file that contains a master record for every geographic feature and the related descriptive position. If required, any descriptive position can be related to coordinates according to ISO 19111 *Referencing by coordinates*. The coordinates may be expressed as point coordinates or as a bounding box for curves or surfaces.

A location type according to ISO 19112 is a territorial unit of the spatial reference system. Examples of location types are an administrative area, town, locality, street, and property. A geographic feature in this context is called a location instance. A geographic feature is listed in the gazetteer and is related to one or more location types; for example, the city hall is a geographic feature that has one record in the gazetteer, and this record contains a position that may be expressed in three ways: as an address, or as a containment in the city, or as a containment in the province (Fig. 15.20).

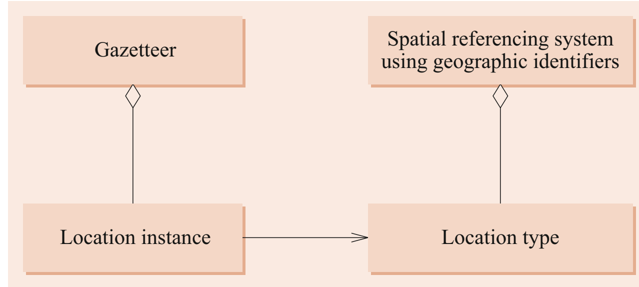


Fig. 15.20 Spatial reference system using geographic identifiers. (Adapted from [41])

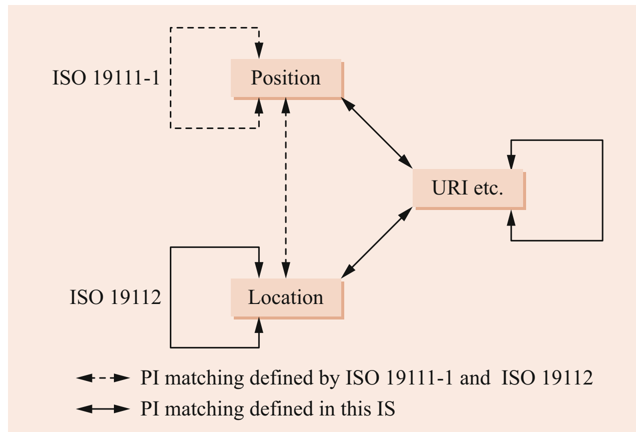


Fig. 15.21 Place identifier matching. (Adapted from [42])

Place Identifier Architecture (ISO 19155)

ISO 19155 specifies an architecture that defines a reference model for an identifier of a place. The concept of *place* includes places not only in the real world but also in the virtual world. These places are identified using coordinate reference systems, geographic identifiers, or virtual world identifiers, such as URIs. In ISO 19155 an identifier of a place is referred to as a place identifier (PI) [42].

The reference model defines a mechanism to match multiple place identifiers to the same place, a data structure, and a set of service interfaces.

If the place is identified with coordinates it is called a *position*, and if the place is identified with geographic identifiers it is called a *location*. Fig. 15.21 illustrates the relation to ISO 19111-1 and ISO 19112.

Place Identifier Architecture – Part 2: Linking (ISO 19155-2)

The ISO 19155-2 *Place Identifier (PI) architecture – Part 2: Place Identifier (PI) linking* extends the ISO 19155 model and defines three mechanisms that define how Place Identifiers can be linked with features or objects in other encodings [43]:

- Id attribute of a GML object (gml:id) as defined in ISO 19136-1:2020
- Universally Unique Identifier (UUID) as defined in IETF RFC 4122
- Uniform Resource Locator (URL) as defined in IETF RFC 1738.

The intended applications include location-based services, linked open data, and robotic assisted services.

Discrete Global Grid Systems (ISO 19170)

In the past, tessellation of a two-dimensional space is performed either on a curved surface, e.g., an ellipsoid, or on a flat surface that is the result of the projection from the curved surface. In both cases, the resulting grid cells do not have equal size. This is a drawback for many subsequent analysis tasks. The series of the standards ISO 19170-x define methods to overcome this disadvantage.

A Discrete Global Grid System (DGGS) [44] is a reference system that uses a hierarchical tessellation of cells to partition the globe providing spatio-temporal referencing by zonal identifiers. DGGSs enable combined analysis of very large, multisource, multiresolution, multidimensional, distributed geospatial data. Discrete Global Grid Systems Specifications – Part 1: Core Reference System and Operations, and Equal Area Earth Reference System [44] specifies the definition of a DGGS composed of a reference system using zonal identifiers with structured geometry and functions providing import, export, and topology query. It provides spatio-temporal classes for geometry, topology, reference systems using zonal identifiers, zonal identifiers, and zones based on ISO 19111 Coordinate RS. It also provides and Equal Area Earth Reference System for an equal area Earth DGGS. It is Part 1 of what is planned to be a four-part standard. Future parts will cover three-dimensional and equal-volume Earth reference system; spatio-temporal Earth reference system; Axis-aligned reference system with all zone edges parallel to the base CRS's axes; specification for a DGGS API; and creation of a register system for DGGS definitions. This ISO standard has an equivalent OGC standard. Figures 15.22 to 15.24 give an example for such an area-preserving tessellation. Figure 15.22 shows a tessellation with the ratio 1 : 9 which preserves the shape of the cell, in this case a square. Figure 15.23 shows that a 1 : 9 tessellation along latitudes and longitudes of the earth does not preserve the area-size. Figure 15.24 shows an example for the method defined in this standard.

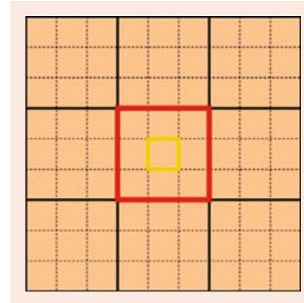


Fig. 15.22 A tessellation with the ratio 1 : 9 which preserves the shape of the cell, a square. (Image source: ISO-TC211_N5496)

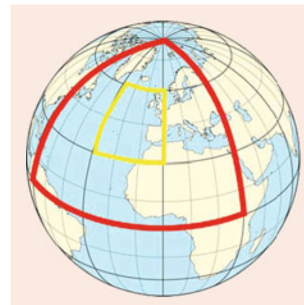


Fig. 15.23 A 1 : 9 tessellation along latitudes and longitudes of the earth does not preserve the area-size. (Image source: ISO-TC211_N5496)

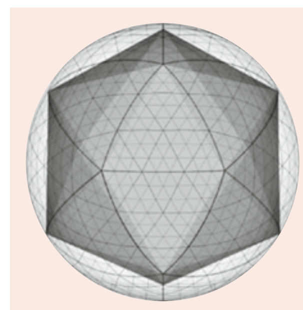


Fig. 15.24 Example for an area-preserving tessellation defined by this standard. (Image source: ISO-TC211_N5496)

Vector Data

ISO 19107 *Spatial schema* is a standard for describing the geometry and topology of geographic information. This standard comprises a comprehensive definition of the geometric elements required to build a geographic dataset.

The standard primarily addresses vector data up to three dimensions. For the two-dimensional case, the standard includes provisions that guarantee seamless coverage of a complete area.

ISO 19107 defines a method to describe the position of a geometric element. A position is named DirectPosition. It includes the coordinates such as x, y, and z depending on the CRS of the dataset and its dimension. This method is used throughout other important standards of the ISO 19100 family, such as ISO 19123 *Schema for coverage geometry and functions*.

This standard does not address graphic portrayal of the geometric elements.

ISO 19136-1 *Geography Markup Language (GML) – Part 1: Fundamentals* defines an almost complete implementation of ISO 19107 in XML.

Section 15.3.6 contains a detailed description of ISO 19107 and discusses the GML standard.

15

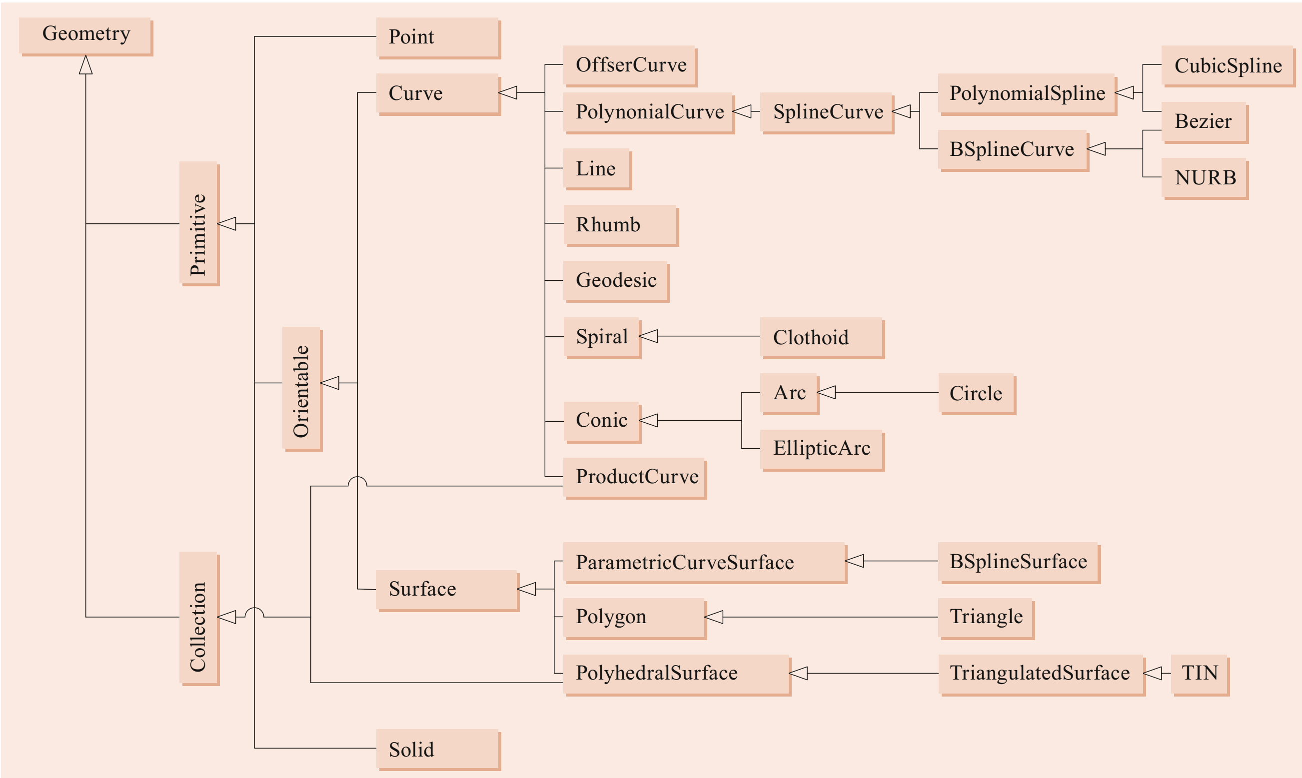


Fig. 15.25 Data model of Geometry (adapted from [45], NURB = Non-Uniform Rational BSpline, TIN = Triangulated Irregular Network)

Spatial Schema (ISO 19107)

ISO 19107 covers a fairly complete domain of vector data and has these characteristics [45]:

- Three-dimensional space, in theory n-dimensional
- Primitives and complexes
- Topological relations
- Based on set theory.

Set theory is the mathematical science of the infinite. It is the study of the properties of sets, abstract objects that pervade the whole of modern mathematics. The language of set theory, in its simplicity, is sufficiently universal to formalize all mathematical concepts, and thus set theory, along with a few other theories, constitutes the true foundations of mathematics [46].

ISO 19107 applies the axioms of set theory throughout the standard. An example is a geometric object that is defined as a set of geometric points.

ISO 19107 has two important design criteria: the boundary criterion and the complexes. The boundary criterion means that high-level elements are composed of a collection of low-level elements; for example, a surface consists of its boundary curves, and the curves are bounded by their start and end points. The complexes, both geometrical and topological, consist of geometries, which do not overlap.

Overview of ISO 19107

Fig. 15.25 shows the logical tree of the important geometry classes of ISO 19107.

General Description of the Geometry Model of ISO 19107

A geometry object (class Geometry) is the most general concept for all objects that a geometrical dataset may consist of. At the top level, a geometry object may be a primitive (class Primitive) or a collection (class Collection). In each case, the vector geometry of a geographic dataset can be completely described based on the fundamental geometries of point, curve, surface, and solid (classes Point, Curve, Surface, and Solid).

What is the difference between primitives and collections? Primitives are the graphic elements that form the complete graphic of a geographic dataset. Primitives exist on their own and have no geometric relations to their neighborhood apart from the common frame of a CRS. A typical map built on primitives and perhaps showing houses, roads, and rivers is primarily designed for visual information.

Collections (class Collection) allow the grouping of geometric elements. For the case of many points that are typically measured by laser scanning or image matching the class PointCloud was included in the ISO 19107.

Curves and surfaces are orientable primitives (class Orientable), while points and solids are not.

A curve is an orientable primitive because it has a defining point sequence that may have a forward or a backward order. This property is important if two or more curves form a closed polygon. A consecutive order of the points is required to correctly define the area as well as to draw a correct line pattern without interruptions along the curve where the individual curves meet. A surface is an orientable primitive because either side may be the upside or downside.

A curve (class Curve) may have many different interpolation types as shown in Fig. 15.25. A surface (class Surface) is defined as a parametric curve surface, a polygon, or a polyhedral surface.

The most sophisticated type of surface is a parametric curve surface (class ParametricCurveSurface). A simple example is a semisphere. For practical reasons, the defining points of a parametric curve surface often lie in a square pattern and thus form a quadrilateral grid. The resulting surface is called a gridded surface.

A simple case of a surface is a closed planar polygon (class Polygon). A typical example is a parcel or a two-dimensional feature in a land use dataset. In ISO 19107, polygons are always planar. This means that all curves belonging to the polygon are part of the same plane. Geometrically, this is always true for triangles only. Quadrangles and other polygons of higher order may have a 3-D shape. Those polygons are not valid polygons according to ISO 19107.

A polyhedral surface (class PolyhedralSurface) is a surface composed of polygons connected along their common boundary curves. If the polygons are triangles, it is called a triangulated surface, with no restriction on how the triangulation is derived. A TIN (class TIN, Triangulated Irregular Network) is a triangulated surface of which the network of the boundary curves satisfies the Delaunay criterion. For each triangle in the network, the circle passing through its vertexes does not contain the vertex of any other triangle in its interior.

The TIN is also covered by ISO 19123 Schema for coverage geometry and functions. ISO 19107 standardizes the description of an existing TIN. ISO 19123 addresses the computation of a TIN and the interpolation of elevations.

Topology

Topology describes the neighborhood relations of geometric data and is mostly used to accelerate computational geometry. Topology is an abstraction of the underlying geometry. Points where two or more curves meet are called nodes. The curves between pairs of nodes are geometrically simplified to straight lines and called edges. The surfaces surrounded by edges are called faces. The term to describe three-dimensional bodies defined by nodes, edges, and faces is "topological solid."

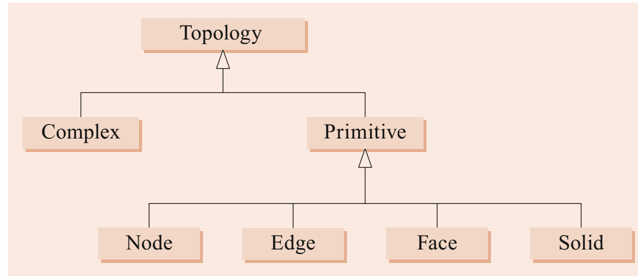


Fig. 15.26 Topology and its subclasses (adapted from [45])

The root class of topology is Topology. It has the subclasses Primitive and Complex. The class Primitive is specialized to its subclasses Node, Edge, Face, and Solid (Fig. 15.26).

A topological complex is a complete network of topological elements (node, edge, face, solid). A topological complex is used to describe that two or more topological networks are disjunct. If they are disjunct, than more than one topological complex exists.

Geography Markup Language (GML) (ISO 19136-1)

ISO 19136-1 *Geography markup language (GML) – Part 1: Fundamentals* standardizes an implementation of the geometry-related standards of the 19100 family, in particular ISO 19107 *Spatial schema* and ISO 19123 *Schema for coverage geometry and functions*. However, ISO 19123 seems to deal more with interfaces, while the coverages in GML are described more from an information viewpoint. GML is an application of XML built on XML schema.

GML is designed for modeling, transport, and storage of geographic data. In a number of predefined schemas, GML provides a rich vocabulary that can be used to create domain-specific GML application schemas. GML serves as a foundation for the geospatial Web and for interoperability of independently developed distributed applications, including location-based services.

GML Schemas

Feature and Feature Collection

A feature is an abstraction of a real-world phenomenon. It is a *geographic* feature if it is associated with a location relative to the Earth. The state of a feature is defined by a set of properties, where each property can be thought of as a {name, type, value} triple.

A feature collection is a collection of features that can itself be regarded as a feature. As a consequence, a feature collection has a feature type and thus may have distinct properties of its own, in addition to the features it contains.

Geographic features in GML include coverages and observation as subtypes.

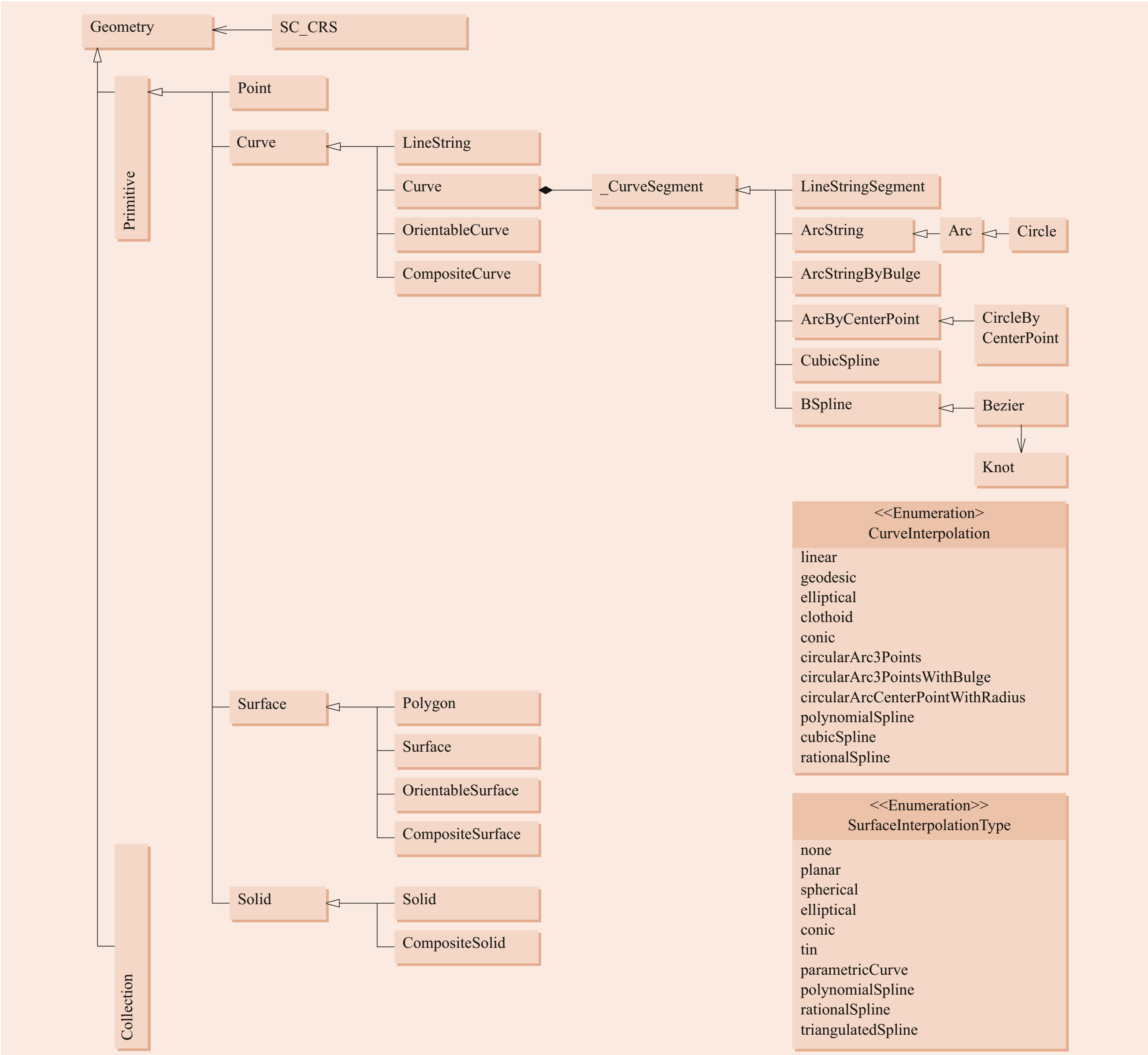


Fig. 15.27 Hierarchy of elements in ISO 19136-1 (GML 3.x)

Geometry

The geometry of a geographic feature describes its location, shape, or extent. The geometry model of GML distinguishes geometric primitives, aggregates, and complexes.

The geometric primitives are the basic elements that are used to form the geometry of a geographic dataset. Primitives are open; that is, a curve does not contain its end points, a surface does not contain its boundary curves, and a solid does not contain its bounding surface.

The geometric aggregates are arbitrary aggregations of geometry elements. They are not assumed to have any additional internal structure and are used to *collect* pieces of geometry of a specified type.

Geometric complexes are closed collections of geometric primitives. This means that they contain their boundaries.

Fig. 15.27 shows the hierarchy of GML geometry types.

Coordinate Reference System

The coordinate reference system (CRS) provides the meaning for location coordinates. A CRS may be associated with any geometry of GML.

Topology

Topology describes the geometric properties of objects that are invariant under continuous deformation; for example,

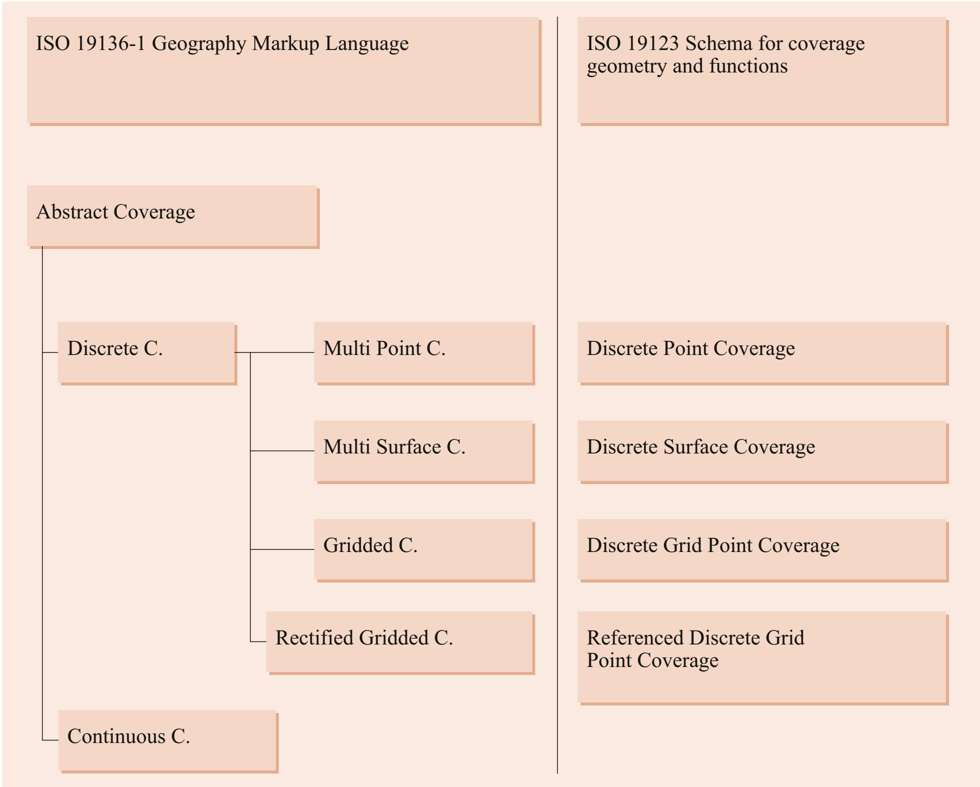


Fig. 15.28 Comparison between the coverage types in ISO 19136-1 and ISO 19123

a square is topologically equivalent to a rectangle and a trapezoid. In geographic modeling, the foremost use of topology is in accelerating computational geometry.

The topology of a dataset is described by the topological primitives nodes, edges, faces, and topological solids. Nodes are topological points where edges meet. Edges are topological lines where faces meet. Faces are topological surfaces where solids meet.

Topological relations are described with boundaries, coboundaries, and directed topological primitives. The topological primitive *edge* is bounded by two directed topological *node* primitives. The topological primitive *edge* is also the coboundary to a pair of nodes.

Temporal Information and Dynamic Features

Time in GML allows the description of the time-dependent status of geographic features and the description of dynamic features, for example, in the domain of location-based services. The definitions of time in GML extend the model of ISO 19108 *Temporal schema*.

Time is measured on two types of scales: interval and ordinal. An interval scale offers a basis for measuring duration. An ordinal scale provides information only about relative position in time, for example, a stratigraphic sequence or the geological time scale.

The default temporal reference system is the Gregorian calendar with coordinated universal time (UTC) [47].

A time instant represents a position in time. It is the equivalent to a point in space. Inexact or *fuzzy* positions may be expressed using the indeterminatePosition attribute that may have values such as *after* or *before*.

A period represents an extent in time. It is the equivalent to a curve in space. It is an interval bounded by beginning and end instants and has a duration.

The status of dynamic features can be described by a snapshot and by a time slice. A snapshot portrays the status of a feature as a whole, whereas a time slice encapsulates the dynamic properties that reflect some change event.

Coverage

Coverages in GML are defined in accordance with ISO 19123 *Schema for coverage geometry and functions* and OGC [48]. However, GML implements only a subset of the functionality defined in the cited sources (Fig. 15.28).

Styling

GML requires strict separation of data and presentation. Therefore, none of the GML data description constructs, such as features and geometries, have built-in capabilities to describe styling information. To simplify the handling of GML, a default styling mechanism was created as a separate model that can be *plugged in* to a GML dataset. The default style schema depends on the W3C synchronized multimedia integration language (SMIL) [49].

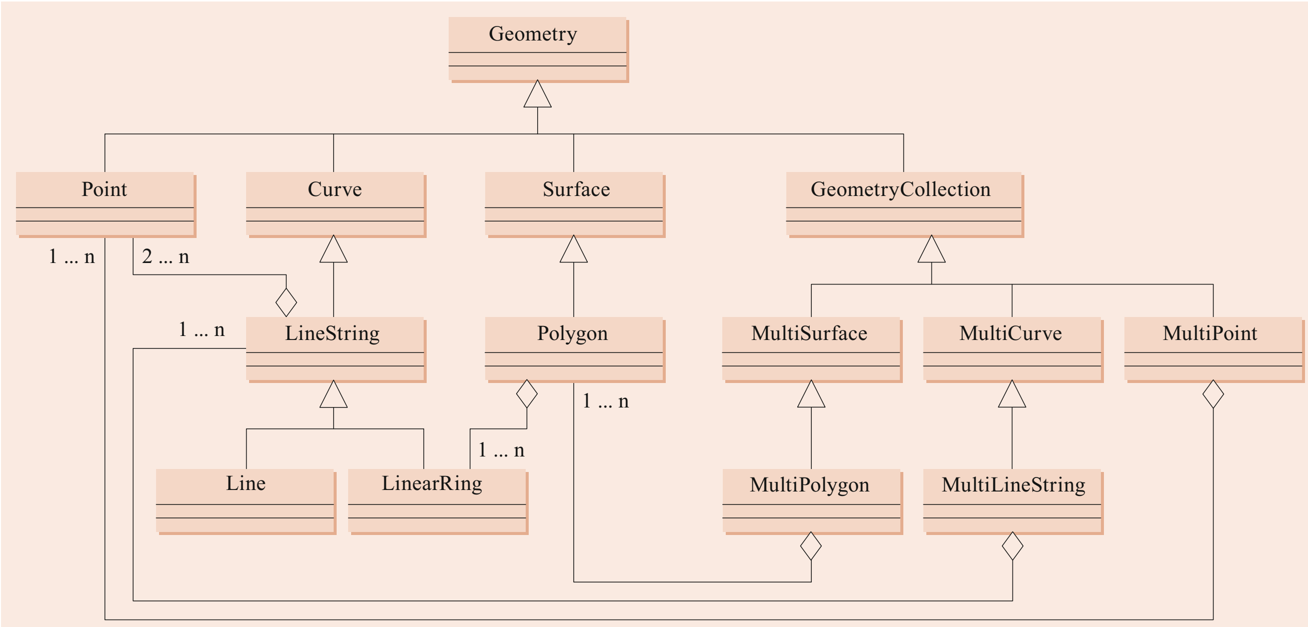


Fig. 15.29 Class-diagram of ISO 19125-1 *Simple feature access – Part 1: Common architecture*

GML – Part 2: Extended Schemas (ISO 19136-2)

ISO 19136-2:2015 *Geography Markup Language (GML) – Part 2: Extended schemas and encoding rules* is an extension to the ISO 19136-1 and contains additions to and simplifications of the GML-standard published with its first edition in 2007, as well as a set of rules to create GML-consistent encodings from other ISO/TC 211 standards [50].

Simplification of the GML standard refers to a reduction of gml tags for several geometry types such as polygons, rectangles, triangles, arcs, and circles. Additions to the GML-standard include support for triangulated irregular networks (tins), linear referencing, and referenceable grids.

The encoding rules define the conversion of code lists and associations modeled in abstract UML class diagrams to GML implementations. Those rules are intended to particularly support applications following ISO 19108 *Temporal schema*, ISO 19123 *Schema for coverage geometry and functions*, and ISO 19148 *Linear referencing*.

Core Profile (ISO 19137)

ISO 19137 *Core profile of the spatial schema* was intended to provide a small subset of the large number of classes of ISO 19107 *Spatial schema* in order to simplify some applications [51]. However, another standard, which implements almost 100% of the spatial schema, became a lot more popular in the meantime, namely ISO 19136-1 *Geography Markup Language (GML) – Part 1: Fundamentals*. According to a systematic review in 2016 this standard is hardly used.

Simple Features

ISO 19125-1 *Simple feature access – Part 1: Common architecture* covers two-dimensional geometries with linear interpolation between vertices [52]. The simple features model consists of the root class geometry and its subclasses Point, Curve, Surface, and GeometryCollection. The class Surface has the subclasses Polygon and PolyhedralSurface, and the class GeometryCollection has the subclasses MultiPoint, MultiCurve, and MultiSurface (Fig. 15.29).

The class Geometry is the equivalent to the class Geometry in ISO 19107. The class GeometryCollection is the equivalent to the class Collection in ISO 19107. The model of ISO 19125-1 does not include complexes, a third dimension, nonlinear curves, or topology.

The simple features gained wide acceptance in spatial database applications because the amount of data required is much less than in the case of ISO 19107. Spatial extension to databases such as Oracle Spatial or PostGIS are applying the simple feature standard. However, more sophisticated applications such as cadastre and cartography tend to use ISO 19107 and ISO 19136-1 *Geography Markup Language (GML) – Part 1: Fundamentals*.

Gap-Analysis: GDF and ISO/TC 211 (ISO 19169)

The geographic data file (GDF) is a file format, which is used to define and exchange digital road databases with a particular emphasis on navigation applications. From the start, GDF was based on similar geospatial concepts as ISO/TC 211 standards. Over the years, both packages took different routes of their development. GDF forms the basis of today's solutions used by TomTom, HERE and other navigational systems while the ISO 19100 family of standards remain the conceptual basis for general geospatial purposes.

With the emergence of increasingly connected and automated road vehicles, there is a need to share geospatial information between the vehicle's navigational and contextual awareness systems and the mapping and road authorities. Therefore, the work underpinning this standard aims to identify the gaps between the two concepts and suggest ways to bridge them.

The ISO/TR 19169:2021 *Gap-analysis: To map and describe the differences between the current GDF and ISO/TC 211 conceptual models to suggest ways harmonize and resolve conflicting issues* has been published as a Technical Report.

Coverage (ISO 19123)

The term "coverage" has different meanings in the geomatics community. It is often a term that is synonymous to a layer that is a thematic subdivision of a dataset. Historically, layers were transparent foils that represented one color of a printed map. Within the ISO 19100 family of standards, coverages have an extended and more abstract meaning. Coverages are a concept considered to describe continuous and discrete spatial and temporal features, thus integrating concepts of the worlds of vector and gridded data. The related standard is ISO 19123 *Schema for coverage geometry and functions* (Fig. 15.30).

The term "coverage", in the sense of the ISO 19100 family, has been adopted from the OGC.

Coverages may be discrete or continuous. An example of a discrete coverage is a map showing cities and their population, where an interpolation between them would make no sense. An example of a continuous coverage is a temperature map that provides a temperature value for any position within the boundaries of a region.

Coverages may be one, two, or three-dimensional. Examples of two-dimensional coverages could be a soil map and a digital elevation model, where the soil type and the elevation, respectively, are handled as attribute values. An example of a three-dimensional coverage is a 3-D grid with values of atmospheric or oceanographic parameters associated with both horizontal position and height (or depth).

Coverages may have a temporal dimension that is the third or fourth dimension in the cases of two or three-dimensional coverages, respectively. An example of a coverage with a temporal dimension is a dataset that contains the recorded daily temperatures of all weather stations of a region over a period of at least several days.

Space and the time are summarized as the spatiotemporal domain. The attributes belong to the attribute domain, which is also called the range, and a coverage may have multiple attributes at each position.

A coverage is a *world of its own* and is viewed as a geographic feature. A coverage includes the operations that are required to use its data, which leads to the perspective that a coverage is a function that relates the spatiotemporal domain to the attribute domain. An example of such operations is a request for an attribute value at a given position.

ISO 19123 standardizes a number of specific coverages. It comprises the Thiessen polygon coverage, the quadrilateral grid coverage, the hexagonal grid coverage, the TIN, and the segmented curve coverage.

Coverage – Part 2: Implementation (ISO 19123-2)

The ISO 19123-2 *Schema for coverage geometry and functions – Part 2: Coverage implementation schema* specifies a concrete and interoperable implementation schema of the ISO 19123. This schema may be encoded in any suitable format such as GML, JSON, GeoTIFF, NetCDF, or GMLJP2. ISO 19123-2 defines a data model allowing different service models to process and deliver coverages such as WFS and WCS.

A coverage is modeled as a specialization of a GML feature (Fig. 15.31).

Raster Data

The topic of imagery includes all geographic data that are stored in an image in the widest sense. The image may display a part of the Earth's surface, or it may contain other fairly evenly spaced data of the Earth's surface or a neighborhood. Examples of imagery-type data are airborne photographs, satellite images, hydrographic soundings, digital elevation models, and coverages according to ISO 19123.

Framework (ISO/TR 19121, ISO 19129)

When the work on ISO standardization of geomatics started, a great variety of raster data formats were in use already. Examples include the Joint Photographic Experts Group (JPEG) and TIFF formats. Therefore, ISO standardization for imagery started with a comprehensive review of industry and other de facto standards.

Enhancements of the Existing ISO 19100 Standards

The following paragraphs explain some of the important additions to existing ISO 19100 standards for the needs of imagery.

A lot of ISO and de facto standards exist for encoding of imagery data. Hardly any standards exist for the storage of associated metadata. ISO/TS 19129 recommends applying the picture coding standards developed by JTC1 SC29, which is responsible for audio, picture, and multimedia information, and JTC1 SC24 *Computer graphics, image processing and environmental data representation*, wherever possible.

Sensors (ISO 19130-1, ISO/TS 19130-2)

Remotely sensed data is an important source for geographic information. Data are called remotely sensed if the sensor has no physical contact with the object measured. For the use of such data in combination with other geographic information, the remotely sensed data must be geometrically

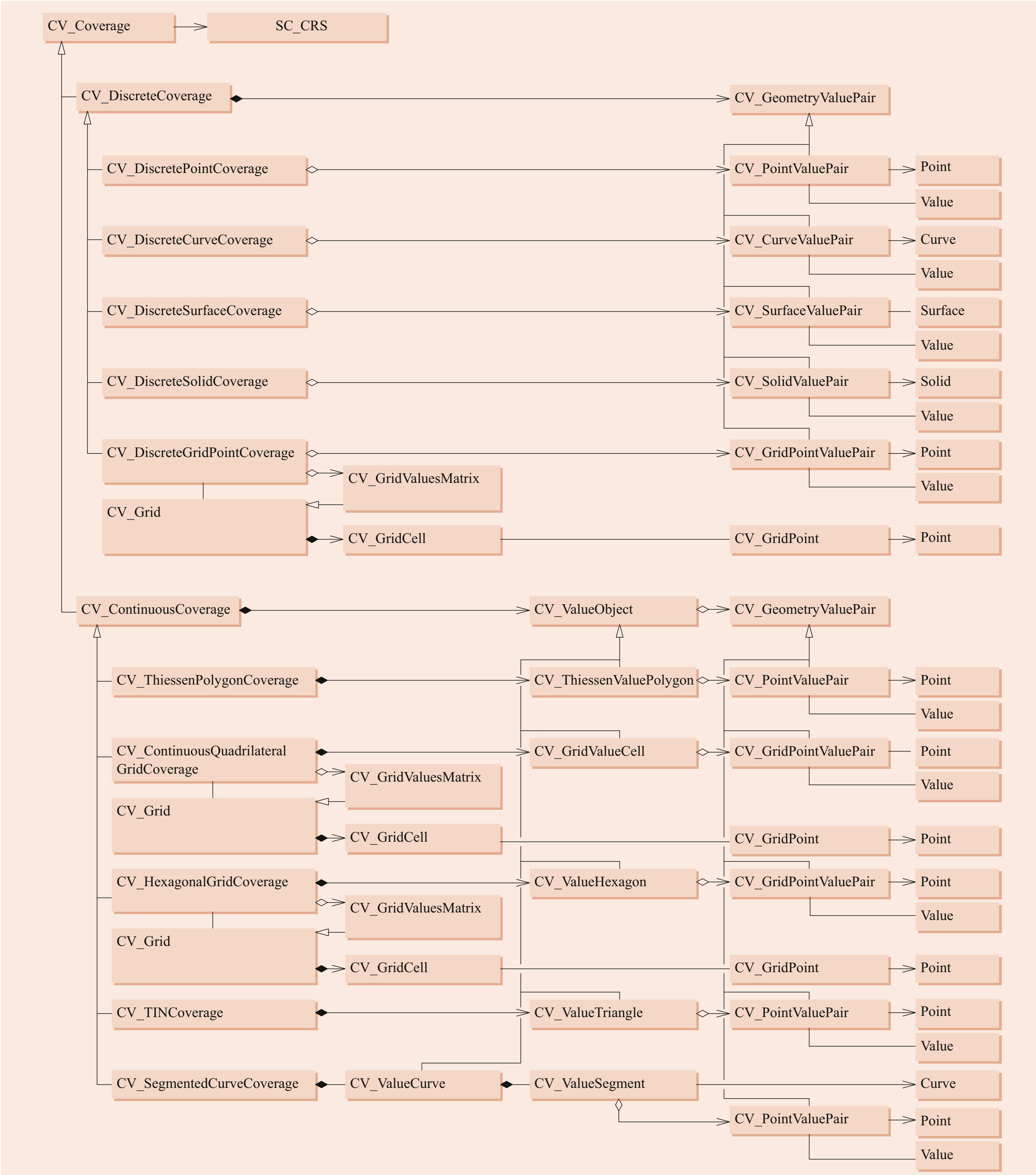


Fig. 15.30 Hierarchy of CV coverage. (Adapted from [53])

referenced to the Earth. ISO 19130-1 *Imagery sensor models for geopositioning – Part 1: Fundamentals* standardizes the metadata for the *geometric* reference of the originally sensed data to locations on the Earth. This reference allows for more precise data retrieval from the imagery. For most applications, it is a prerequisite for appropriate use of the imagery.

For reasons of efficient organization of the standard's development, the topic has been split into two documents. ISO 19130-1 addresses physical sensor models (line and

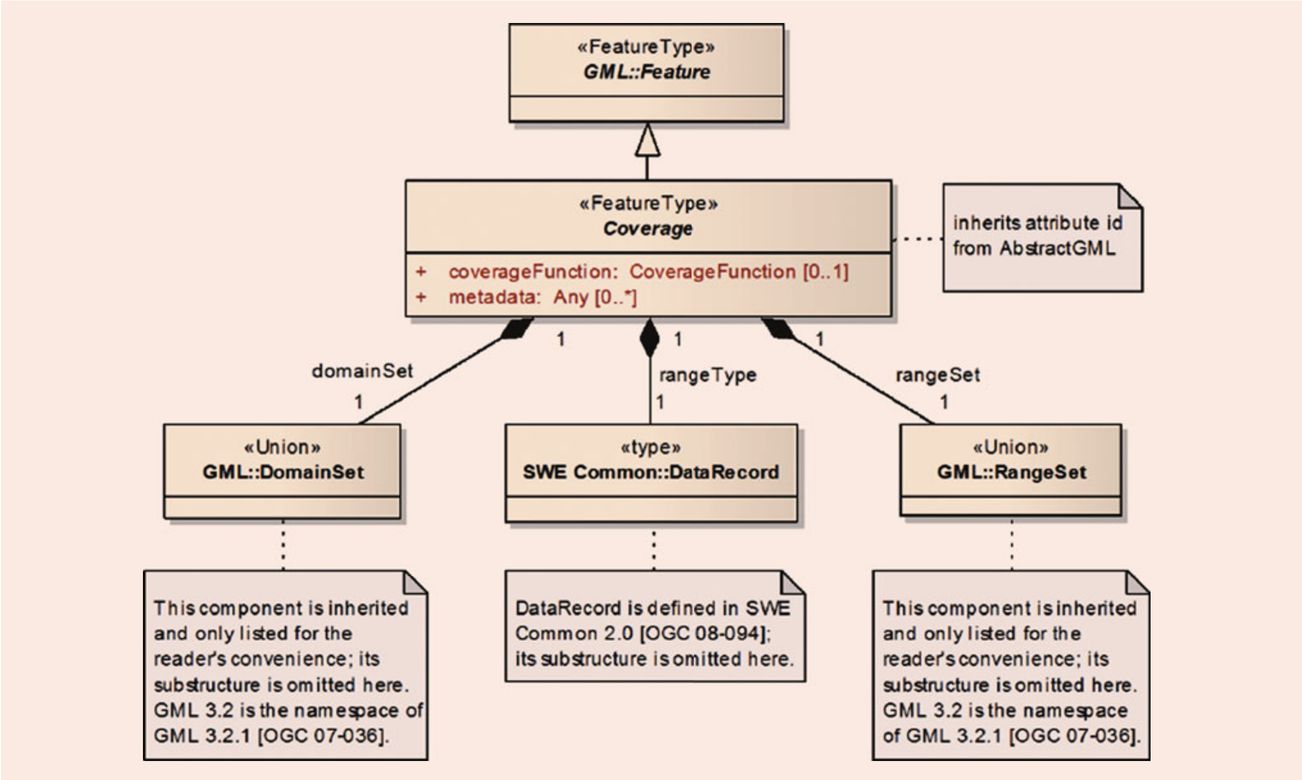


Fig. 15.31 Coverage as a subclass of GML::Feature

matrix cameras), true replacement models, and correspondence models. ISO/TS 19130-2 adds synthetic aperture radar (SAR)/interferometric SAR (InSAR), light detection and ranging (lidar), and sound navigation and ranging (sonar). The series ISO/TS 19159-1/2/3 *Calibration and validation of remote sensing imagery sensors* completes the set with a calibration and validation standard for the sensors involved.

Model Classification of ISO 19130-1

From a conceptual view, the standard comprises sensor models for all remote-sensing sensors [54]. At present, the sensor models comprise the physical sensor model, the true replacement model, and the correspondence model.

The Physical Sensor Model

The physical sensor model stands for the rigorous photogrammetric approach that conceptually tries to describe any geometric offset from the model of the central perspective. The parameter sets are typically named interior orientation (calibrated focal length, principal point of autocollimation, distortion) and exterior orientation (position and attitude of the sensor, and eventually its dynamics).

ISO 19130-1 defines the use of ground control points as a georeferencing method on its own and puts it under the physical sensor model. The relation between the original image and the Earth's surface is defined by a set of common points. Each point consists of two or three ground point coordinates (x, y, and optional z) and the two related sensor data coordinates (row and column). It is not specified which functional relation between image and Earth shall be used, and thus the type and values of the coefficients remain undefined.

The physical sensor model comprises frame cameras, pushbroom scanners, and whiskbroom scanners.

A *frame camera* produces a matrix of image pixels. Apart from some rare exceptions, the camera is mounted on an aircraft (Fig. 15.32).

A *pushbroom sensor* takes the data along a scanline at one moment (Fig. 15.33).

Most scanning linear arrays and pushbrooms are flown on satellite platforms. As the orbit is smooth compared with the ground resolution, the resulting image may be georeferenced by robust transformations (sensor reference, rigorous model) or by functional relations, such as polynomial transformations (image reference). If a scanning linear array-type sensor is operated from an airplane, line-by-line rectification is necessary to georeference the data (Chaps. 9 and 10).

A *whiskbroom scanner*-type sensor scans the terrain within the field of view under the flight track. This sensor scans along one scan line using a rotating mirror or similar device (Fig. 15.34).

15

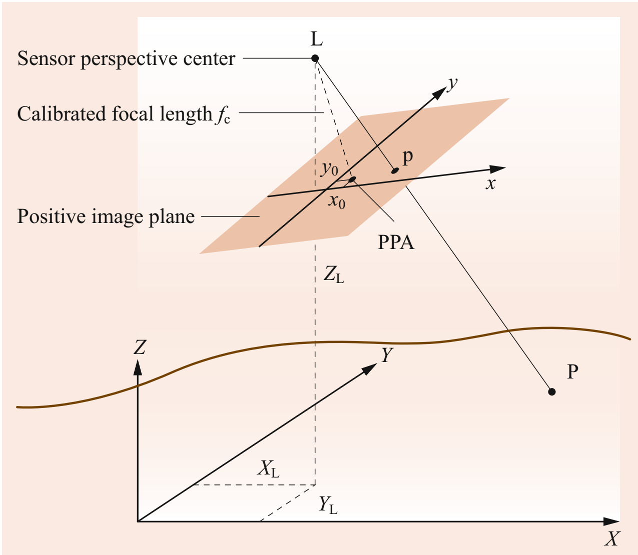


Fig. 15.32 Frame camera image and its exterior orientation. (Adapted from [54]) (f_c – calibrated focal length, given in (mm), PPA – principal point of autocollimation, given in (mm) in the image coordinate system, x_0, y_0 – coordinates of principal point of autocollimation, L – sensor perspective center (projection center), given in (m) in the object coordinate system, p – image point, P – object point, x, y – image coordinate system, X, Y, Z – object coordinate system)

True Replacement Model

True replacement models are produced using physical sensor models. The equations that describe the sensor and its relationship to the Earth's coordinate reference system are replaced by a set of equations that directly describe the relationship between image coordinates and Earth's coordinates. This method originated from military applications in order to disguise sensor characteristics.

The specification allows different interpolation methods for finding the ground point coordinates within the grid provided, as shown in Fig. 15.35.

Grid Interpolation

In this method, geolocation is derived by interpolation in an evenly spaced grid. Similarly to the ground control points method, each grid point consists of three ground point coordinates and the two related sensor data coordinates. ISO 19130-1 does not define the interpolation method to be used between the grid points.

Polynomials

The polynomials method presents the sensor data coordinate (row or line l and column or sample s) as a function of ground point coordinates (X, Y, and optional Z) in the form of a polynomial. For the polynomials, l and s may be written as

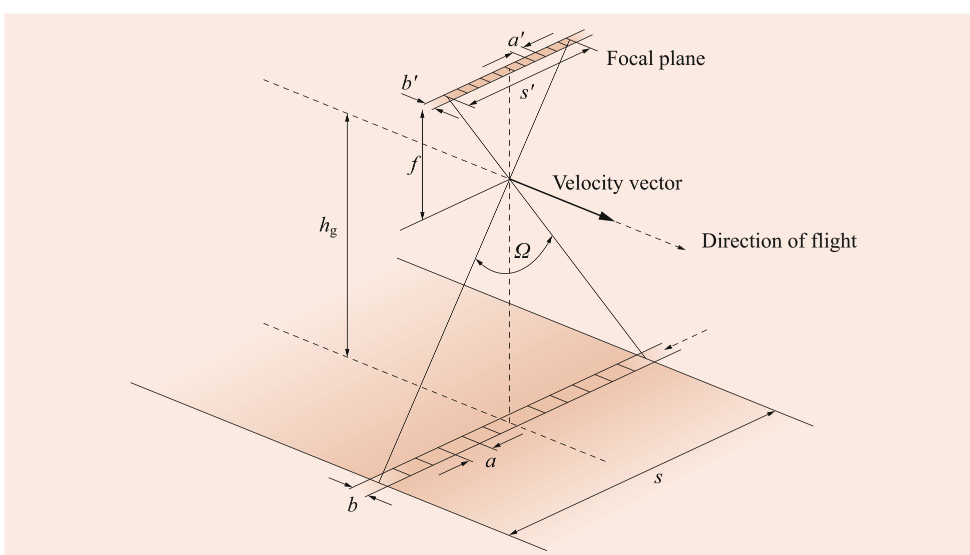
$$l = \sum_{k,l,m}^{n} a_{klm} X^k Y^l Z^m \,,$$

$$s = \sum_{k,l,m}^{n} b_{klm} X^k Y^l Z^m \,.$$

Ratios of Polynomials

Rather than using a single polynomial, this geolocation method uses separate ratios of two polynomial functions of latitude, longitude, and height to compute image row and column. It is often called the rational functions model, and the polynomial coefficients are often called rational polynomial coefficients (RPC) data. An image is divided into segments, and different polynomial ratios are used for the different sections. Each polynomial has 20 terms, although the coefficients of some polynomial terms are often zero. In the polynomial functions, the three spatial reference coordinates and two image coordinates are each offset and scaled, so that their range over an image segment is from -1.0 to $+1.0$.

Fig. 15.33 Optical layout of a pushbroom sensor. (Adapted from [54]) (h_g – flying height above ground, f – focal length, $a' \times b'$ – image pixel size, s' – length of scan array, Ω – field of view (FOV), $a \times b$ – object (ground) pixel size, s – swath width)



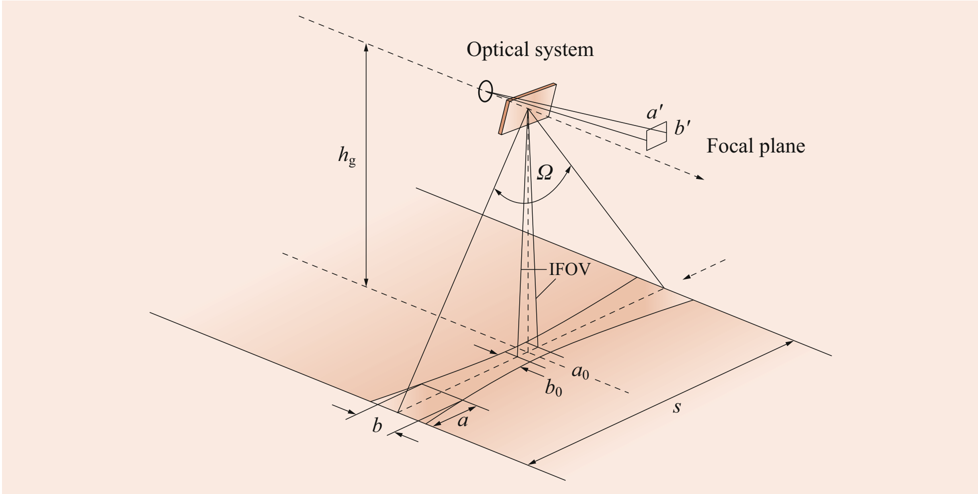


Fig. 15.34 Optical layout of a whiskbroom scanner. (Adapted from [54]) (h_g – flying height above ground, f – focal length, $a' \times b'$ – image pixel size, s' – length of scan array, IFOV – instantaneous field of view, $a \times b$ – object (ground) pixel size, s – swath width)

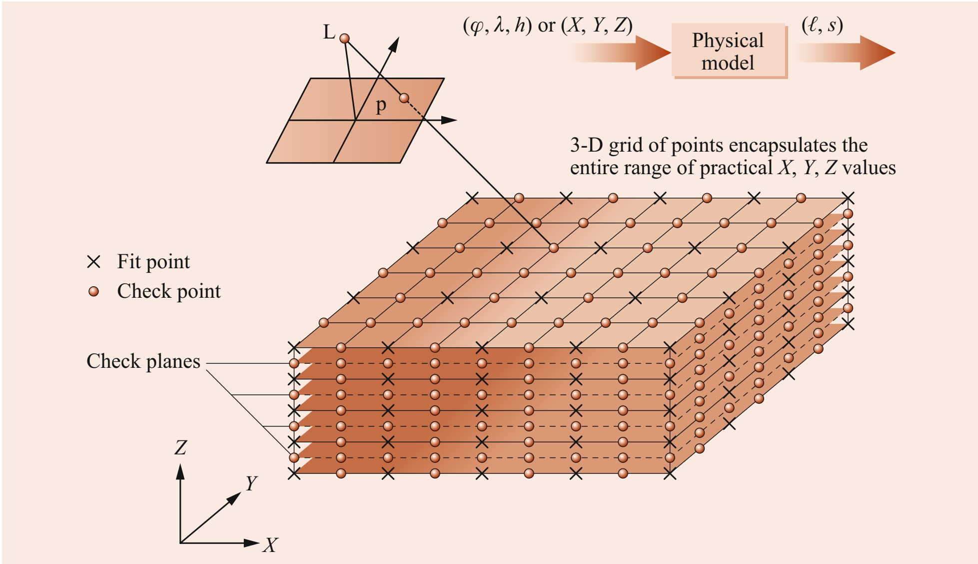


Fig. 15.35 True replacement model generation of points. (Adapted from [54])

For each image segment, the defined ratios of polynomials have the form

$$l = \frac{P_1(X,Y,Z)}{P_2(X,Y,Z)}\,, \quad s = \frac{P_3(X,Y,Z)}{P_4(X,Y,Z)}\,,$$

where l is the image line coordinate, s the image sample coordinate, and X, Y, Z the object ground coordinates.

The polynomials P have the form

$$P = \sum_{k,l,m}^{n} a_{klm} X^k Y^l Z^m\,,$$

where a_{klm} are the polynomial coefficients.

Direct Linear Transform

When the ground coordinates are in a Cartesian CRS, such as geocentric or local space rectangular coordinates, the RPC model reduces to the special case containing only 11 coefficients of the form

$$l = \frac{a_0 + a_1 X + a_2 Y + a_3 Z}{1 + c_1 X + c_2 Y + c_3 Z}\,,$$
$$s = \frac{b_0 + b_1 X + b_2 Y + b_3 Z}{1 + c_1 X + c_2 Y + c_3 Z}\,,$$

where l is a line coordinate, s is a sample coordinate, X, Y, Z are object space coordinates, a_0, a_1, a_2, a_3 are the coefficients of the numerator of the rational polynomial that produces a line coordinate, b_0, b_1, b_2, b_3 are the coefficients of the numerator of the rational polynomial that produces a row coordinate, and c_1, c_2, c_3 are the coefficients of the denominator of the rational polynomials.

These equations are known as the direct linear transform (DLT).

Correspondence Model

The correspondence model assembles all those approaches that make no use of a physical sensor model and are, thus, generally less accurate.

The correspondence model is a *functional fit model* that relates an image to the Earth by simply using a set of common points. The points define a functional relation, such as the polynomials explained above, which is used to warp the original image towards the geometry of the Earth's surface. The functional fit model establishes a so-called *image reference* between the remotely sensed data and the ground. Again, the model does not require any knowledge of the sensor.

Model in ISO/TS 19130-2

ISO/TS 19130-2 *Imagery sensor models for geopositioning – Part 2: SAR (synthetic aperture radar)/InSAR, lidar, and sonar* adds the sensor types that are quoted in the title to the previously defined ISO 19130-1 [55]. The standard also incorporates a metadata model for aerial triangulation.

SAR, InSAR

Interferometric SAR, or InSAR, is the application of interferometry to SAR images. An interferogram, the interferometric overlay of two SAR images, is produced by taking amplitude (intensity) and phase of both images into account. InSAR allows for three-dimensional measurement of a surface and may be applied to detect surface deformation or mapping topography.

Synthetic aperture radar (SAR)- and *interferometric SAR* (InSAR)-type sensors are operated from satellite and aircraft platforms (Figs. 15.36 and 15.37). The sensor emits radiation in the microwave band and records the time and phase of the reflection. The system works under any light conditions (i.e., day or night). As microwave frequencies can penetrate clouds, the system's operation is not restricted to good weather conditions.

Lidar

A *lidar* sensor emits laser pulses and records the time until a reflection is received. It is mostly operated from an aircraft in order to determine the shape of the Earth's surface. An airborne lidar is also called *laser scanning*. Some systems include seabed surveys in shallow water.

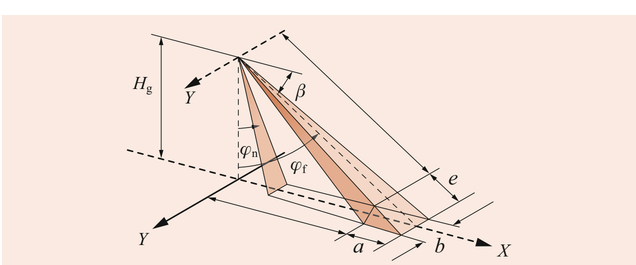


Fig. 15.36 Geometry of side-looking synthetic aperture radar (SAR). (Adapted from [55]) (H_g – flying height, β – depression angle, φ_n – near-edge incidence angle, φ_f – far-edge incidence angle, a – ground range resolution (x-direction), b – azimuth resolution (y-direction), e – slant range resolution)

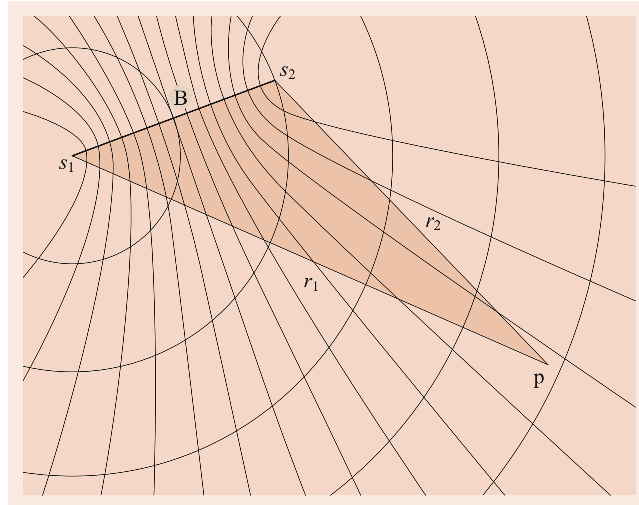


Fig. 15.37 Interferometric point positioning by range and phase difference. B is the base between two independent antennas. The *circles* correspond to *equirange* lines to sensor s_1, and the hyperbolas to *equidifference of range* lines to both sensors. The flight direction is vertical to the drawing plane. (Adapted from [55])

ISO/TS 19130-2 defines four basic types of lidar systems:

1. Range finders, which are used to measure the distance from the lidar sensor to a solid or hard target.
2. Differential absorption lidar (DIAL), which is used to measure chemical concentrations (such as ozone, water vapor, and pollutants) in the atmosphere.
3. Doppler lidar, which is used to measure the velocity of a target.
4. Multiple-receiver lidar, which allows accurate location of a target in three dimensions by adding multiple receivers at different locations and triangulating the results.

Sonar

Hydrographic sonar is used to measure the depth of the seabed and emits sound, as well as measuring the time until a reflection is received (Fig. 15.38).

ISO/TS 19130-2 defines four basic types of sonars:

1. A single-beam echo sounder system produces one narrow sonar beam directly beneath the transducer and receives a return from the closest point at which it intersects the seabed.
2. Swath (multibeam or interferometric) sonar systems have a single transducer or pair of transducers that transmit a fan-shaped signal perpendicular to the ship's direction of travel.
3. Sweep or boom systems are characterized by several transducers mounted on a boom, which is then operated parallel to the water's surface and orthogonally to the vessel's direction of travel.

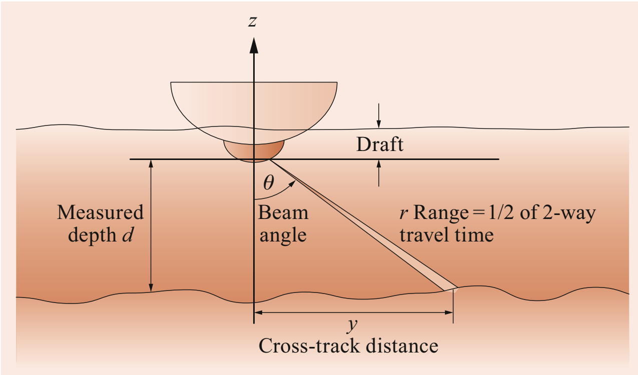


Fig. 15.38 Position and depth calculation for a hydrographic sonar. (Adapted from [55])

4. Side-scan sonar systems generally have two transducers mounted transversely in a towfish.

Sensors – Part 3: Implementation (ISO/TS 19130-3)

While the ISO 19130-1 and the ISO/TS 19130-2 define abstract models, the future ISO/TS 19130-3 *Imagery sensor models for geopositioning – Part 3: Implementation schema* will define an XML-schema implementation based on OGC's sensor model language (SensorML). The schemas shall address all sensors covered in parts 1 and 2, i.e., frame sensor, pushbroom, and whiskbroom sensors, synthetic aperture radar, lidar, and sonar. The development has not been completed at the time of publication of this book.

Calibration and Validation (ISO/TS 19159-1/2/3)

ISO/TS 19159-1/2/3 *Calibration and validation of remote sensing imagery sensors* addresses the calibration and validation of those remote-sensing sensors where a georeferencing method is defined in ISO 19130-1 and ISO/TS 19130-2. Those sensors operate on airborne and spaceborne platforms and deliver imagery data. Examples are frame cameras, line cameras, lidar, and SAR.

The term "calibration" refers to geometry and radiometry and includes the instrument calibration in a laboratory, as well as in-situ calibration methods, such as field measurements. The term "validation" denotes a general approach to evaluating the quality of a process or a dataset. This specification defines validation for specific use cases, such as the validation of calibration parameters that were found some time ago.

Essentially, ISO/TS 19159-1/2/3 are metadata definitions that extend ISO 19115-1 and ISO 19115-2 and a description of calibration and validation processes.

Calibration and Validation – Part 1: Optical Sensors (ISO/TS 19159-1)

The first part of this standard defines about 150 metadata elements that allow the documentation of the calibration results of a remote-sensing and a photogrammetric camera [56]. The parameters are grouped in UML classes that are dedicated to the general environment, e.g., photo flight, radiation, to the calibration facility, in-flight or laboratory, and to sensor-specific properties, e.g., geometry, radiometry, and the detector elements. The standard supports the distortion-models of Brown, Fraser, SMAC (simultaneous multiframe analytical calibration), Ebner, and Jacobsen.

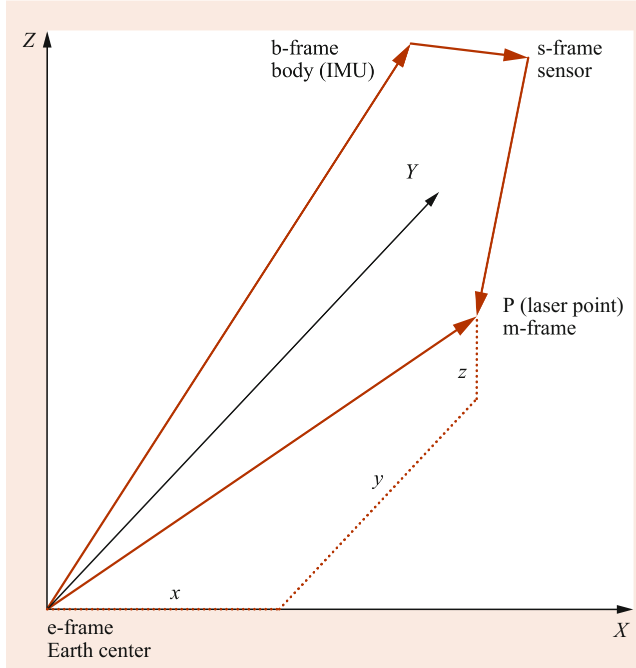


Fig. 15.39 Transformation chain of an airborne lidar measurement

Calibration and Validation – Part 2: Lidar (ISO/TS 19159-2)

The second part of this standard defines about 50 metadata elements related to the calibration of an airborne lidar sensor [57]. The parameters cover the description of the sensor, several attributes to describe its calibration, the rotation angles of the transformation chain from scanner to ground, and a description of the testfield.

Fig. 15.39 and Table 15.5 illustrate the transformation chain (e-frame – b-frame – s-frame – m-frame (P)).

Calibration and Validation – Part 3: SAR/InSAR (ISO/TS 19159-3)

This part of the standard, number 3, covers the calibration of SAR (synthetic aperture radar) and InSAR (interferometric SAR) [58]. It standardizes about 130 metadata elements that are related to seven topics related to the SAR sensor: antenna system, radar system, antenna phase center, signal processing, atmospheric propagation and Earth motion, SAR calibration field, and SAR validation. The InSAR case adds some parameters that are specific to a pair of receivers.

Table 15.5 Frames defined in the ISO 19159-2

ID	Description
s-frame	Sensor: frame of the laser sensor, defined by the principal axes of an optical instrument
b-frame	Body: frame materialized by the triad of accelerometers within an IMU
l-frame	Local-level: this frame is tangent to the global ellipsoid (normally WGS84), with the orthogonal components usually defined as N-orth (x), E-ast (y) and D-own (z)
e-frame	ECEF frame (Earth-centered Earth-fixed). The origin is the geocenter of the earth, x-axis points towards the Greenwich meridian and the z-axis is the mean direction of the earth rotation axis
m-frame	Mapping frame. Cartesian frame with E-ast (x), N-orth (y) and U-p (z) component

Though this is not mentioned in the standard's title, a section is dedicated to PolSAR, the polarimetric SAR. This method exploits the property of coherent electro-magnetic radiation to oscillate in exactly one plane in space, resulting in the possibility to generate four different observations at the same time with an advantage for many remote-sensing tasks.

Content Components – Part 1: Content Model (ISO/TS 19163-1)

This Technical Specification, the ISO/TS 19163-1, aims at the encoding of imagery data independent of one of the privately developed image formats such as GeoTIFF, JPEG 2000, HDF, HDF-EOS, BIIF, and NetCDF [59]. Based on the frameworks defined in ISO 19101-2:2018 *Reference model – Part 2: Imagery and gridded* data and ISO 19123:2005 *Schema for coverage geometry and functions*, it specifies categories of imagery and gridded data based on thematic and spatial attributes, as well as sensor types.

Content Components – Part 2: Implementation (ISO/TS 19163-2)

This Technical Specification, the ISO/TS 19163-2, defines a GML-based implementation schema of ISO 19163-1. It also provides an implementation schema to bind a coverage structure as defined in ISO 19123-2.

Visualization: Portrayal (ISO 19117)

Originally, the graphic presentation of geographic information was strictly the domain of cartography but has now also become an important section of geomatics. ISO 19117 *Portrayal* defines a schema to create graphic output for datasets and metadata of the ISO 19100 family of standards [60]. The scope of ISO 19117 does not include standardization of cartographic symbols, standardization of symbol graphics such as Scalable Vector Graphics (SVG), portrayal services such as Web Map Server (WMS), dynamic maps, map generalization, or the third dimension as used in simulations.

Chap. 13 of this handbook gives a comprehensive view on the science of cartography.

Portrayal Catalogue

According to ISO 19117, the cartographic symbolization is kept separate from the feature types of the dataset. The definition of the cartographic representation for a feature is stored in a portrayal catalogue. Essentially, the catalogue is a reference list that relates each feature code that is used to identify different feature types to an individual cartographic portrayal.

The portrayal catalogue contains portrayal symbols and portrayal functions. The symbols are meant to hold their basic geometry and graphics. The functions modify those properties to adapt to the mapping scale and other display parameters (Fig. 15.40).

Conditional Portrayal Functions

The standard distinguishes between two types of functions: conditional and context.

The conditional functions can test for feature attributes, geometry, and other properties of the feature; for instance, this function determines which symbol will be selected for a given feature. The function may be simple, such as a *black*

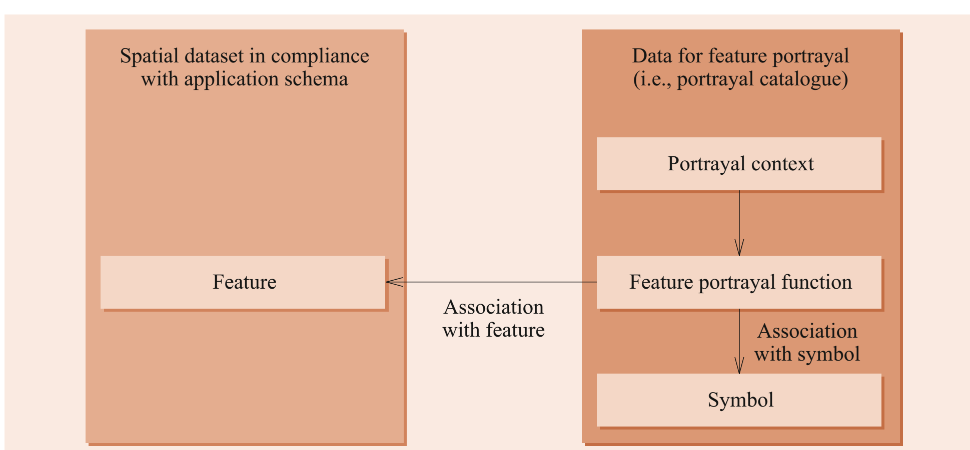


Fig. 15.40 Overview of portrayal. (Adapted from [60])

solid line to portray a local road, or sophisticated, such as a *double dashed red line* to portray a major road that carries 10,000 or more vehicles daily and that is maintained by the provincial government.

Examples of topographic feature types are streams, fields, and elevation points. A portrayal catalogue contains one cartographic representation for each feature type. For these examples, it may contain a blue line for streams, a green fill-area symbol for fields, and a brown point for elevation points.

Context Portrayal Functions

The context function can test for external factors, such as display scale and viewing conditions, which are relevant, for instance, in mobile applications. An example is a car navigation system that will always display the map with the driving direction up.

Default Representation, Priority, External Functions

The portrayal specification contains all the attributes and operations required to derive the graphic representation for the given feature type according to the applicable function. The catalogue will always specify a default representation if the function search for a given feature type fails. More than one portrayal catalogue may be defined for one dataset to allow for different types of maps.

Dense graphics may result in uncontrolled overlay of elements. A priority attribute allows defining the top element that hides the others in the case of multiple elements at the same position.

To adapt the graphics to any given symbolization catalogue, external functions may be applied to the individual functions or to the whole dataset.

15.3.6 Terminology

This section discusses the ISO/TC 211 approach to maintaining a consistent terminology within the technical committee and the strategy to align with the terminology from other domains.

Terminology (ISO 19104)

ISO 19104 *Terminology* provides guidelines for collection and maintenance of terminology in the field of geomatics. The main concept for setting up terminology demands that the same term be used for the same concept throughout the whole family of standards. ISO 19104 also lays down the guidelines for maintenance of the terminology repository [61].

Each standard addresses its segment of geographic information and uses the most appropriate terms found there. The ISO terminology standard guarantees the consistency of all terms involved through the specification of a terminological record and the description of principles for definition writing.

In practice, consistent terminology among all ISO 19100 standards has not yet been achieved. This is due to the fact that some of the standards were developed outside the ISO/TC 211 environment. Examples are ISO 19123 *Schema for coverage geometry and functions*, which was proposed by the OGC, and ISO 19136-1 *GML*, which was originally developed by a private company.

Much of the terminology has now been harmonized. The terms that still have to be kept with different meanings are marked in the list of definitions.

As ISO standards must be reviewed every 3 to 5 years, a family of standards undergoes minor, but continuous, change. To guarantee current and efficient management of common terms, a terminology repository has been established. The database covers all ISO 19100 terms, showing their definitions and providing notice where multiple concepts for the same term apply. The database is updated continuously. In this Handbook, the section *Terms and Definitions of the ISO 19100 Standards* is an outcome of it as of 2021.

Cross Domain Vocabularies (ISO 19146)

Typically, geographic information lies at the point of intersection of many kinds of disciplines. They range from the local cadastre to world meteorological maps, or from car navigation systems to health monitoring. Consequently, terminologies from very distinctive domains meet, which often leads to overlap of terms or their underlying concepts [62].

ISO 19146 *Cross-domain vocabularies* defines a methodology to overcome this terminology problem that is caused by adopting the technical vocabularies from different industry-focused geospatial communities. Words can have several meanings depending on the context, while concepts can be referenced by several words, each communicating different connotations and levels of emphasis.

ISO 19146 does not intend to define an ontology or taxonomy for geographic information. The ontology, i.e., the explicit specification of the concept of geographic information, is the topic of the ISO 19150-x group of standards. The taxonomy, i.e., the science and practice of classification, has not yet been addressed by ISO/TC 211, apart from a brief section in ISO 19144-1 *Classification systems – Part 1: Classification system structure*.

ISO 19146 sets seven principles for cross mapping of vocabularies:

1. The terminology is to be consolidated rather than proliferated. The purpose of the cross mapping of vocabulary is to standardize the association of specific terms with specific concepts. However, it should not be used as a mechanism for permanently entrenching unnecessary duplication in terminology conventions.
2. Vocabulary cross mapping shall provide a thesaurus, not a taxonomy or ontology. The standard assumes that the

developers of the relevant standards have already established the subject area vocabularies.

3. A stable reference vocabulary, maintained by a recognized standards body, shall be adopted for all cross-mapping undertakings involving a particular discipline. In the case of ISO/TC 211, the basis is its multilingual glossary of terms. The terms in this Handbook's section *Terms and Definitions of the ISO 19100 Standards* are based on this glossary.
4. Cross mapping shall proceed as a collaborative venture. During the process, each community of interest should be acknowledged as the ultimate authority regarding the correct use and interpretation of its terms and definitions.
5. Cross mapping shall not circumvent established processes. In spite of this principle, cross mapping may trigger other processes within the collaborating organizations to deprecate terms or to improve concept system structures.
6. Cross mapping should be recognized through publication in a register. ISO 19135-1 *Procedures for item registration – Part 1: Fundamentals* shall be applied to define the register.
7. Cross mapping should accommodate continuous change. The cross mapping of concepts should be periodically reviewed to identify and accommodate any changes [62].

15.3.7 Metadata

Metadata are defined as "information about a resource". While the majority of metadata can be found in ISO 19115-1, several other ISO 19100 standards also define metadata. ISO 19115-2 is an extension for acquisition, processing, and imagery. ISO/TS 19115-3 defines an implementation of parts 1 and 2. Other standards, such as ISO 19159-1/2/3 (calibration), are modeled as specilizations of ISO 19115-1 and, thus, add further metadata.

Metadata – Part 1: Fundamentals (ISO 19115-1)

ISO 19115-1 *Metadata – Part 1: Fundamentals* describes the metadata for documenting geographic information. It provides information about the identification, extent, quality, spatial and temporal schema, spatial reference, and distribution of digital geographic data and services. This standard is applicable for describing geographic datasets, dataset series, and individual geographic features and feature properties. This standard can be used for cataloguing datasets, clearinghouse activities, and the full description of datasets and services.

The standard contains the complete listing of all metadata elements of the ISO 19100 family, with a short explanation of each; however the detailed description of most elements can only be found in the individual standards.

Metadata is a prerequisite to reuse geographic datasets. In a heterogeneous computing environment that has a variety of available data sources, as well as a great number of different applications for the data, the metadata provide guidance to find the most appropriate dataset for a certain application. The metadata are capable of locating, evaluating, extracting, and employing the required datasets.

The current version of ISO 19115-1 refers primarily to vector data (Fig. 15.41). ISO 19115-2 *Metadata – Part 2: Extensions for acquisition and processing* is an add-on for imagery-type data (raster) and other applications.

The Concept of Metadata

In the past, metadata was supplied as additional but separate pieces of information to a dataset. However, the concept of ISO 19115-1 currently considers metadata as an integral part of the data that travels with data when copied, moved, renamed, or exported, and never gets lost. Metadata is implemented using standard technology such as XML, and large GIS manufacturers have made the ISO 19115-1 metadata part of their system software.

Hierarchy of Metadata

Metadata standards existed before the work on ISO 19115-1 started, as the development of ISO 19115-1 intended to make it a superset of the existing metadata standards in the field of geographic information. A basic approach to metadata is the Dublin Core [64] (see later). The design of ISO 19115-1 follows ISO/IEC 11179, which addresses the specification and standardization of data elements. Work is currently continuing in an attempt to align the existing metadata standards to ISO 19115-1.

The Dublin Core metadata element set is a basic standard for resource description in the IT business and, in particular, in library management. This standard was developed by the Dublin Core Metadata Initiative in cooperation with the National Information Standards Organization (NISO), USA, and approved by ANSI as ANSI/NISO Z39.85 – 2001. The Dublin core defines only 15 metadata elements, because simplicity lowers the cost of creating metadata and promotes interoperability. On the other hand, simplicity does not accommodate the semantic and functional richness supported by complex metadata schemes. The design of the Dublin core mitigates this loss by encouraging the use of richer metadata schemes in combination with Dublin core. The name Dublin core refers to Dublin, Ohio, USA, where the first workshop of the initiative was held in 1995. For further reading, see Chap. 14.

Metadata – Part 2: Extensions for Acquisition and Processing (ISO 19115-2)

ISO 19115-2 *Metadata – Part 2: Extensions for acquisition and processing* extends ISO 19115-1 by defining the

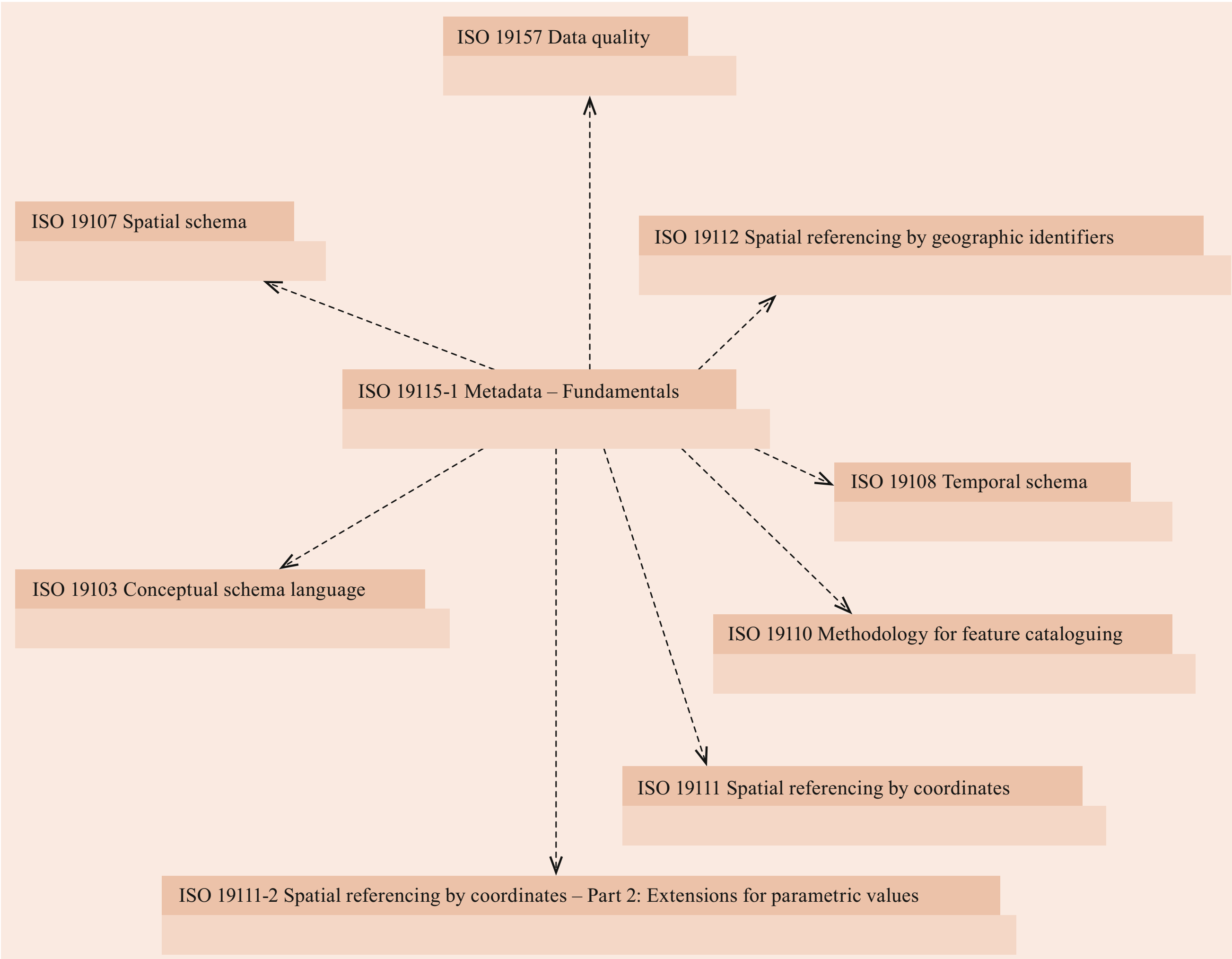


Fig. 15.41 UML package diagram and dependencies from *Metadata fundamentals* (after [63])

schema required for enhanced description of the acquisition and processing of geographic information, including imagery [65]. The standard also includes the properties of measuring systems and the numerical methods and computational procedures used to derive geographic information from the data acquired.

Metadata – Part 3: Implementation (ISO/TS 19115-3)

The ISO/TS 19115-3 *Metadata – Part 3: XML schema implementation for fundamental concepts* provides an integrated implementation of parts 1 and 2 following the rules defined in ISO 19118 and ISO/TS 19139-1.

ISO/TS 19115-3 defines the following artefacts:

- A set of XML schema conforming to conceptual model elements defined ISO 19115-1, ISO 19115-2, and ISO/TS 19139-1.
- A set of Schematron rules (ISO/IEC 19757-3) that fill a gap of the XML schema functionality as they implement validation constraints defined in the ISO 19115-1 and ISO 19115-2 UML models, which are not validated by the XML schema.
- An Extensible Stylesheet Language Transformation (XSLT) for transforming ISO 19115-1:2014 metadata encoded using the former ISO/TS 19139-1 XML schema and ISO 19115-2:2019 metadata encoded using the former ISO/TS 19139-2 XML schema into an equivalent document that is valid against the XML schema defined in the new ISO/TS 19115-3.

The resulting XML schemas are algorithmically derived from the normative UML models of the ISO/TC 211 standards. The resulting schemas are predictable, since UML classes, attributes, associations, etc., are encoded following consistent rules and patterns.

Metadata – Implementation – Part 1: Encoding rules (ISO/TS 19139-1)

The previous version, ISO/TS 19139:2007 *Metadata – XML schema implementation,* defined rules and an XML encoding for the metadata elements defined in ISO 19115:2001 *Metadata*. The new version, ISO 19139-1, only contains the XML encoding rules. The XML encoding of ISO 19115-1 is covered by ISO/TS 19115-3 as described above.

Data Product Specifications (ISO 19131)

ISO 19131 *Data product specification* promotes the application of the ISO 19100 standards family for practical uses. A data product is a dataset containing data with a spatial relation to the Earth. A data product specification is a formal document that defines the details of the data product for production, end-users, and other purposes. It is a precise technical description of the data product in terms of the requirements that it will or may fulfill. Therefore, the purpose of the standard is to provide practical help in the creation of such specifications.

A good guideline for the content of a data product specification is listed in a normative annex to the ISO 19131 *Data product specification* standard. It helps to take all details into account without necessarily thoroughly knowing the structure of the whole ISO 19100 family.

15.3.8 Quality

The value of geographic data is directly related to its quality. The views on quality differ among the various user communities; for example, cadastral applications require positional accuracy within a few centimeters, whereas a nautical chart requires a positional accuracy of only a few meters. Geographic datasets are being increasingly shared, interchanged, and used for purposes other than the producer's intended uses. The purpose of describing the quality of geographic data is to facilitate the selection of the geographic dataset best suited to application needs or requirements.

The topic of quality is spread across three ISO 19100 standards, data quality (ISO 19157), its implementation (ISO/TS 19157-2), and quality assurance (ISO/TS 19158).

Data Quality (ISO 19157)

ISO 19157 *Data quality* defines the *data quality elements* that contain *quantitative* quality information. ISO 19157 defines six types of data quality:

- *Completeness*. Presence and absence of features, their attributes and relationships. Negative example: missing road data in a remote part of the province.
- *Logical consistency*. Degree of adherence to logical rules of data structure, attribution, and relationships. Example: the application schema distinguishes between public and private buildings. The dataset distinguishes between low buildings and high-rises.
- *Spatial accuracy*. Accuracy of the position of features in relation to the Earth. Example: the absolute point accuracy is 10 cm (diagonal).
- *Temporal quality*. Quality of the temporal attributes and temporal relationships of features. Example: the date of the data compilation was August 2010.
- *Thematic accuracy*. Accuracy of quantitative attributes and the correctness of nonquantitative attributes, as well as the classification of features and their relationships. Example: areas have been classified according to remotely sensed imagery as meadows although they were swamps.
- *Usability*. Degree of adherence to a specific set of data quality requirements. Example: a product specification fully applies for the intended purpose. However, the data are 12 years old.

Data quality may be reported using metadata or a stand-alone report. Data quality shall be reported as metadata in compliance with ISO 19157 and ISO 19115-1. To provide more details than reported as metadata, a standalone report may additionally be created. Its structure is not defined.

Data Quality – Part 2: Implementation (ISO/TS 19157-2)

ISO/TS 19157-2 *Data quality – Part 2: XML schema implementation* defines an XML schema implementation of ISO 19157:2013, the data quality related concepts from ISO 19115-2 utilizing encoding rules from ISO 19118, ISO/TS 19139-1, and the implementation approach from ISO/TS 19115-3.

Quality Assurance (ISO/TS 19158)

ISO/TS 19158 *Quality assurance of data supply* defines a framework for the producer and customer in their production relationship. It bridges the gap between the quality-management systems as defined in the ISO 9000 family of standards and the technically-oriented quality standards of the ISO 19100 family, such as ISO 19157 *Data quality*.

Through the application of ISO/TS 19158 there are opportunities for a better understanding of requirements by all involved in production and update, especially within multiple-producer environments. Additional benefits may be reduced data throughput time, reduced rework, improved data quality, and increased confidence within a relationship, leading to lower costs for both supplier and organization.

The existence of ISO/TS 19158 does not impose its application, thus releasing smaller companies from potentially unnecessary organizational overheads. The question of whether a demanded quality level has been reached is answered by second-party accreditation. This is performed by the customer regarding the quality of the supplied data without the involvement of a third-party accreditation body [66].

15.3.9 Temporal schema (ISO 19108)

As far as geomatics is concerned, temporal characteristics are standardized in ISO 19108 *Temporal schema*. Many geographic applications require a time stamp related to the physical reality, and therefore the standard deals with the valid time, which is the time when the event occurred in the abstracted reality. The standard does not address the transaction time, when the data becomes available in a database [67].

In many applications, time is handled as a metadata element only. More advanced applications require time as a further dimension. This enables modeling of the behavior of a feature as a function of time, such as a satellite's position along its orbit. The time dimension may be combined with spatial and parametric coordinate reference systems to form a compound CRS [33, 34].

The standard distinguishes between the geometry of time and the topology of time. The geometry of time has the four major classes:

1. The instant: a certain time.
2. The period: the time elapsed between two instances.
3. The order: the sequence of instances.
4. The relative position: temporal relation of earlier and later instances.

The topology of time describes the temporal connectivity between two or more occurrences. If two or more occurrences take place at one instant, that instant is called a node. More precisely, the period between the occurrences is smaller than the resolution of time. If two or more occurrences take place simultaneously during the period, that period is called an edge.

More temporal aspects are a part of ISO 19136-1 *Geography Markup Language (GML) – Part 1: Fundamentals*, discussed in Sect. 15.3.6.

15.3.10 Catalogue

The world of catalogues-standards for geographic information has not evolved in a straightforward manner. It started with feature catalogue (ISO 19110), followed by the more general approach of a registry (ISO 19135-1). NATO demanded the feature dictionary (ISO 19126) as one of the first technical specifications. The procedures for item registration (ISO/TS 19135-2:2012) was completed as an implementation standard.

Registration – Part 1: Fundamentals (ISO 19135-1)

As the number of applications for digital geographic information is growing rapidly, a centralized registration is a means of promoting compatibility and avoiding duplicate efforts. ISO 19135-1 *Procedures for item registration* specifies procedures for registration of items of geographical information [68].

A registration is the assignment of an unambiguous name to an object. When bound to a set of files, the objects form a register. The registry is the information system that maintains one or many registers. The registration authority keeps the registry.

The parties submitting their application to the registration authority are called sponsors. An application may include a profile of the ISO 19100 series of standards, a feature catalogue, or other specifications for geographic information. A typical case for a registry is the registration of coordinate reference systems (CRS). Their definition hardly changes over time, and it is essential that coordinate transformations in different applications use exactly the same parameter sets. With the help of a registry, the parameters can be retrieved online.

The standard allows more than one registration authority, but they must be approved by the ISO. To allow the user of geographic information to keep an overview of the authorities involved, their number shall be kept as small as possible. The registration authority keeps the registry as an Internet database.

The registry allows dynamic update of the registered objects. Although it is not essential, the items may be specified in an international standard. ISO 19145 *Registry of representations of geographic point location* is an example of such a standard, based on ISO 19135-1. See Chap. 18 for more details.

Registration – Part 2: Implementation (ISO/TS 19135-2)

ISO/TS 19135-2 *Procedures for item registration – Part 2: XML schema implementation* has been requested by the user community, e.g., DGIWG, because current implementations come up with their own XML schema implementations for ISO 19135-1, which will inevitably lead to interoperability issues. ISO/TS 19135-2 aims at harmonizing such encodings [69].

Feature Dictionaries (ISO 19126)

ISO 19126 *Feature concept dictionaries and registers* may be considered as a link between feature catalogues and registers, ISO 19110 and ISO 19135-1, respectively. ISO 19126 demands in a fairly abstract way the *basic definitions and related information about a set of concepts* that shall be used to describe the features that will become the elements of a feature catalogue. ISO 19126 also describes how those feature concepts can be maintained in a standardized registry [70].

The Defence Geospatial Information Working Group (DGIWG) was one of the major promoters of this standardization project. Consequently, an annex of this standard describes the DGIWG feature data dictionary (DFDD), a profile of ISO

19126, as an example of implementation of the feature concept dictionary schema. The *DGIWG feature and attribute data registry* can be accessed online without restrictions.

Feature Catalogues (ISO 19110)

The name of ISO 19110 is *Methodology for feature cataloguing*, with a feature being the fundamental unit of geographic information [17]. Feature catalogues define the types of features, their operations, attributes, and associations represented in geographic data. The details of the modeling of a feature and its relationships to other features are explained in *Rules for application schema (ISO 19109)*.

A feature catalogue forms a repository for a set of definitions to classify significant real-world phenomena to a Universe of Discourse. The catalogue provides a means for organizing the data that represent these phenomena into categories to ensure the resulting information is as unambiguous, comprehensive, and useful as possible.

A feature catalogue has many purposes. The most important ones are listed here:

- A feature catalogue may be sufficient for many applications, thus enabling cost reduction because one catalogue can be applied for many purposes.
- A feature catalogue should present an application-oriented abstraction of the real world represented and in a form readily understandable and accessible to users of the data.
- Often the feature types in different systems have equal or similar names, such as private houses and private buildings. A feature catalogue may serve to clarify where the classification differs.

ISO 19110 provides a template for the organization of feature catalogue information. This template comprises sections for each feature type and the associated attributive information.

15.3.11 Web Services

There are currently many means of communication being used, but the Internet is admittedly the most important. If we use the Internet, we do not usually think of the software running in the background. We take it for granted. Our interface to the net is a browser. A great variety of maps or other visual representations of geodata are available on the Internet.

The most important Web mapping standards have been approved for publication as ISO-standards, i.e., Web map server (ISO 19128), Web feature service 2.0 (ISO 19142), filter encoding as a part of the Web feature service (ISO 19143), and the latest, Geospatial API for features, Core (ISO 19168-1).

Aside from standardizing web service applications ISO/TC 211 is standardizing Web APIs as building blocks which can be assembled by developers to access and manipulate geospatial information. Web API being standardized will be organized by resource type. Initially, ISO/TC 211 has standardized building blocks dealing with geographic features. This new series of standards is titled "Geospatial API for Features" with its part 1 "Core".

Web Map Server (ISO 19128)

A Web map server produces maps of georeferenced data. In this context, a map is a visual representation of geodata, but it is not the data themselves. ISO 19128 *Web map server interface* is a specification that standardizes the way maps are requested by clients via the Internet and the way that servers describe their data holdings. ISO 19128 is a Web interface specification for mapping data [71].

Originally, the ISO standard Web Map Server interface was a development of the Open Geospatial Consortium (OGC) [72]. See Chap. 16 for more detail.

Operations

The ISO standard Web Map Server interface defines three operations, the first two of which are required of every Web map server. All requests are sent from the client to the server:

- GetCapabilities (required): obtain service-level metadata that is a machine-readable (and human-readable) description of the Web Map Server's information content and acceptable request parameters.
- GetMap (required): obtain a map image for which the geospatial and dimensional parameters are well defined.
- GetFeatureInfo (optional): ask for information about particular features shown on a map.

Processing

A standard Web browser can ask a Web map server to perform these operations by simply submitting requests in the form of uniform resource locators (URLs). When invoking the operation GetMap, a Web Map Server client can specify the information to be shown on the map. Some of the information that can be specified would include the number of layers, the *styles* of those layers, what portion of the Earth is to be mapped (a *bounding box*), the projected or CRS to be used, the desired output format, the output size (width and height), and the background transparency and color. When invoking GetFeatureInfo the client indicates what map is being queried and which location on the map is of interest.

According to the philosophy of the Internet, no main server is responsible for collecting the data requested by the client. The display data commonly reside only at the client's computer. A particular Web Map Server provider in a distributed Web Map Server network only needs to be the steward of its own data collection.

Because each Web Map Server is independent, a Web Map Server must be able to provide a machine-readable description of its capabilities. The *service metadata* enables clients to formulate valid requests and enables the construction of searchable catalogues that can direct clients to particular Web Map Servers.

A Web Map Server may optionally allow the GetFeatureInfo operation. If it does, its maps are said to be *queryable*, and a client can request information about features on the map by adding a position around which nearby features are sought.

Web Feature Service (ISO 19142)

The purpose of ISO 19142 is to establish a Web mapping interface that supports streaming geographic information, primarily GML, across the Internet [73]. WFS also supports the modification of features across the Internet; this additional capability is referred to as transactional or WFS-T for short. The interface includes raster as well as vector data and works on the feature and feature property levels. This essentially implements a client–server structure with which a geographic dataset can be fully maintained, i.e., where features can be created, modified, or deleted.

The method of noting information by key–value pairs (KVP) is often used in information technology. An example is the GetCapabilities request of the Web Feature Service, such as `http://hostname:port/path?service=WFS&request=GetCapabilities`, where, e.g., *request* is the key, and GetCapabilities is the value.

ISO 19142 defines an XML and a KVP encoding of a system-neutral syntax for 11 operations. Because of complexity, the definition of a KVP encoding is not feasible for two of the operations. See Chap. 16 for more detail.

GetCapabilities

The GetCapabilities operation generates a metadata document describing a WFS service provided by a server similar to the Web map server. At the beginning of a session, this document is requested by the client.

DescribeFeatureType

The DescribeFeatureType operation returns a schema description of feature types offered by a web feature service (WFS). The schema descriptions define how the WFS expects feature instances to be encoded on input (via insert, update, and replace actions) and how feature instances shall be encoded on output (in response to a GetPropertyValue, GetFeature, or GetFeatureWithLock operation). This operation is also invoked at the beginning of a session to align the client's capabilities with those of the server.

GetPropertyValue

The GetPropertyValue operation allows the retrieval of the value of a feature property from the data store for a set of features identified by a query expression.

GetFeature

The GetFeature operation returns a selection of features from a data store to the client. The document contains zero or more feature instances that satisfy the query expressions specified in the request. The standard representation of features uses GML.

LockFeature

Locking of features is a necessary operation as soon as two or more editing processes work on the same geospatial database, as otherwise it may happen that a feature is worked on by two editors at the same time, which would lead to undetermined results.

The purpose of the LockFeature operation is to expose *a long-term feature locking* mechanism to ensure consistency. The lock is considered long term because network latency would make feature locks last relatively longer than native commercial database locks.

GetFeatureWithLock

The GetFeatureWithLock operation provides functions similar to the GetFeature operation except that, in response to a GetFeatureWithLock operation, a WFS will not only generate a response document similar to that of the GetFeature operation but will also lock the features in the result set, presumably to update the features in a subsequent transaction operation.

Stored Queries

A stored query is a named, persistent, parameterized query that can be invoked numerous times with different values bound to the query parameters. Thus, this simplifies repeated queries of the same type. Like all WFS queries, it returns a number of features that satisfy the stored query for the specified parameter values. Stored queries allow the simplification of complex temporal and spatial queries.

ISO 19142 defines four operations related to stored queries: CreateStoredQuery, DropStoredQuery, ListStoredQueries, and DescribeStoredQueries

Transaction

Using the transaction operation, clients can *create*, *modify*, *replace*, and *delete* features in the WFS data store. The transaction is, therefore, the most comprehensive operation of the WFS. However, this operation is optional.

Filter Encoding (ISO 19143)

The term "filter" as used in ISO 19143 *Filter encoding* denotes an encoding of predicates that are typically used in query operations to specify how data instances in a source dataset should be filtered to produce a result set. Originally, the content of this standard was a part of OGC's WFS specification but has been separated from ISO 19142 *Web Feature Service* because that type of filtering is not limited to the application of ISO 19142.

Like ISO 19142, the ISO 19143 also defines an XML and keyword–value pair (KVP) encoding of a system-neutral syntax for query expressions and an XML encoding for the listed predicates; for example, an XML-encoded query could be transformed into an SQL SELECT … FROM … WHERE … ORDER BY … statement to fetch data stored in a SQL-based relational database. Similarly, the same XML-encoded query expression could be transformed into an XQuery expression in order to retrieve data from an XML document. These queries could be expressed with a KVP encoding in a similar manner.

ISO 19143 defines only the XML encoding for the predicates [74]:

1. Logical predicates: and, or, and not
2. Comparison predicates: equal to, not equal to, less than, less than or equal to, greater than, greater than or equal to, like, is null, and between
3. Spatial predicates: equal, disjoint, touches, within, overlaps, crosses, intersects, contains, within a specified distance, beyond a specified distance, and BBOX (bounding box)
4. Temporal predicates: after, before, begins, begun by, contains, during, ends, equals, meets, met by, overlaps, and overlapped by
5. A predicate to test whether the identifier of an object matches the specified value.

Geospatial API for Features – Part 1: Core (ISO 19168-1)

The development of the OGC-webservices of which some have been also standardized by ISO/TC 211 took place in the years before the turn of the century. At that time the general information technology did not offer many tools necessary for building a web mapping environment. As a consequence, most of the basic libraries were developed with XML-technology within OGC's project. This text refers in particular to the Web Map Server (ISO 19128) and the Web Feature Service (ISO 19142 in conjunction with ISO 19143).

Today, extensive libraries with IT standard technology are freely available which allow a redesign of web mapping, now simpler in structure, broader in functionality, and well understood in the IT-community. The new version of the standard shall remove the dependency on XML and XML-schema and is intended to leverage modern web functionality using WebAPIs using the OpenAPI specification https://github.com/OAI/OpenAPI-Specification.

Geospatial API for Features will be a multi-part standard with a separation between core requirements and more advanced capabilities. Geospatial API for Features – Part 1: Core [75] specifies discovery and query operations implemented using HTTP GET methods restricted to features represented in the WGS 84 coordinate reference system with axis order longitude/latitude. Support for additional reference systems, creating and modifying features and dealing with more complex data stores will be standardized in future parts. The standards development work is being performed jointly with OGC.

The Geospatial Features – Part 1: Core offers support for four encodings: HTML, GeoJSON, GML level 0, and GML level 2. The decision on the most appropriate encoding for a given application may be based on its complexity. GeoJSON is mainly an implementation of the ISO 19125-1 *Simple feature*. Thus, several features such as non-linear curve-interpolation are not available. In contrast to that, GML offers more complex requirements than GeoJSON.

15.3.12 Observations and Measurements (ISO 19156)

Sensor web enablement (SWE), an activity of the OGC, was one important reason for the development of ISO 19156 *Observations and measurements*. SWE is concerned with establishing interfaces and protocols that will create a sensor Web through which applications and services will be able to access sensors of all types, and observations generated by them, over the Internet.

ISO 19156 provides a very general approach to the nature of observations as they are defined as an act associated with a discrete time instant or period through which a result – a number, a term, or another symbol – is assigned to a phenomenon. The phenomenon is a property of the feature being observed [76].

The observation may involve the application of a specified procedure, such as a sensor, an instrument, an algorithm, or a process chain. This procedure may be applied in situ, remotely, or ex situ (such as in a laboratory) with respect to the sampling location.

The respective ISO/IEC guide [77] relates the term "measurement" to any operation that aims at determining a value of a quantity. However, ISO 19156 uses the term "observation" for the general concept. "Measurement" is reserved for cases where the result is a numeric quantity.

Metrology is the science of measurements and has its origins in mechanical engineering. The measurement tasks in spatial sciences differ slightly from typical tasks in metrology in that they are related to space and time, and – more importantly – often only allow for indirect measurements or include mass data.

A typical example for an indirect measurement is the analysis of a remotely sensed imagery that is applied for land use mapping. This remotely sensed image is an intermediate product from which the ultimate feature of interest, the map, is derived.

The observation model of ISO 19156 emphasizes the semantics of the feature of interest and its properties. Its viewpoint may be characterized as *data-user*-centric. This contrasts with sensor-oriented models, which are often more technical and take a process-centric or provider-centric viewpoint.

15.3.13 Digital Rights

ISO/TC 211 has addressed digital rights of spatial data through standards for a reference model (ISO 19153) and a rights expression language (ISO 19149).

Reference Model (ISO 19153)

ISO 19153 is a reference model for Digital Rights Management (DRM) functionality for geospatial resources (GeoDRM) [78]. This ISO standard evolved from an OGC abstract specification.

In the digital world, due to the nature of digital resources and commerce, most digital entities are sold in a manner different from traditional trade. When a user acquires an application, he or she actually acquires the right to use a copy of the application. Possession does not equate with ownership, and a system of software and resource licensing has grown up in the digital world that ensures the following types of things:

- The user may legitimately act upon a resource if he or she has a corresponding license for that act.
- The owner should maintain the resource, fixing errors (*bug fix*) and ensuring a guaranteed level of functionality.
- Optionally, the user may be asked to pay the owner of the resource based upon agreed criteria, whether that is a one-time fee, a per-machine fee, a usage fee, or some other mechanism stated in the legal contract or license between user and owner.
- The user agrees to protect the owner's rights based on the agreement. This usually means he or she cannot backward-engineer code or resource, nor redistribute the resource without proper permission.
- The owner agrees to maintain the resource and allow reasonable access to users for any fixes that may be required. Again, the extent or degree of maintenance is stated in the user agreement.
- To create and support a large-scale, open market in geospatial resources, this type of protection is needed to ensure that a fair value for work (investment) ethic can be guaranteed so that suppliers can be sure of fair return on individual sales, and users can be sure of fair value for purchases of uses of resources [78].

Roadmap

To allow the definition of interfaces and responsibilities, seven possible function packages are defined by ISO 19153. Those packages may be considered as a roadmap to understand this standard:

- *Rights model*. The rights model defines the basis for developing a geospecific rights expression language, as well as other specifications necessary to establish a GeoDRM-enabled spatial data infrastructure (SDI).
- *Rights expression language*. This package provides the capabilities to express usage rights in the form of a machine-readable and machine-processable representation. However, the definition of this language is not part of this standard.
- *Encryption*. This package includes the required functionality to protect a GeoDRM-enabled SDI against fraud. First, encryption enables the protection of a license so that it cannot be modified by an adversary in order to obtain additional rights. Second, encryption is also useful to protect the digital geographic content against unlicensed use. Because security and trust are not geospecific, ISO 19153 does not define a specific standardization for this type of data.
- *Trust*. An example of a mechanism to establish trust between entities in a service-oriented architecture (SOA) is adding authenticity information on the digital content that is been exchanged between the partners. This mechanism, typically called a digital signature, is not geospecific and, therefore, is not a relevant topic of ISO 19153.
- *License verification*. License verification has to occur before the rights of the license can be enforced. Because document authentication is not geospecific, it is not a topic for ISO 19153.
- *Enforcement and authorization*. The rights expressed in a GeoLicense need to be enforced. The acceptance or denial decision for a particular request (with its associated licenses) is based on the authorization decision, as derived by the authorization engine. Because enforcement and authorization is geospecific, the appropriate standardization will be based on ISO 19153.
- *Authentication*. The basic requirement for trust, license verification, and enforcement/authorization is proof of identity, as provided by the functionality of this package. Authentication is not geospecific and, thus, not part of ISO 19153 [78].

Rights Expression Language (ISO 19149)
When selling data or other digital resources it is known, or at least anticipated, that those products are devaluated because of a far wider application than that covered by the granted licenses. The multimedia industry has taken the lead in solving this problem by creating a general model for digital rights protection. This model includes a rights expression language, specifically defined in ISO/IEC 21000-5 *Multimedia framework (MPEG-21) – Part 5: Rights expression language*, or ISO REL for short, which in conjunction with Digital Rights Management (DRM) systems can protect the value of data and still allow it to be distributed subject to a system of licensing, trust, and enforcement.

ISO 19149 extends the ISO REL to encompass the concerns of holders of geographic data and service resources to equally ensure their protection.

15.3.14 Ontology

The word ontology is derived from the Greek language and, in its original philosophical sense, means the study of being, becoming, existence, reality. The term has been adopted by information technology, in particular by artificial intelligence, however, with a slightly different meaning.

Proposed by Tim Berners-Lee, the founder of the World-Wide Web, the www was extended towards the semantic Web. This extension followed the idea of changing the Web from an unstructured store of documents to an open database with interlinked packages of content. Such a package consists of data, information, and documents enhanced by embedded semantics, in essence, metadata. A semantically enhanced area of the Internet can for practical reasons only be established for a clearly bordered topic. Modeling calls such a limited area a "universe of discourse". An ontology is a program that structures the entities of a specific universe of discourse and associates them to human-readable text, for instance database records.

A number of standards have formalized ontologies. An example is OWL, the Web ontology language. Sometimes, the concept of the semantic Web and other Web extensions is considered rather ambitious, because, to some extent, the success of the www originates from its open and hardly standardized structure.

ISO 19150 *Ontology* covers ontologies as far as spatial data are concerned [79]. Conceptually, ISO 19150 consists of six parts:

1. ISO 19150-1 *Ontology – Part 1: Framework*
2. ISO 19150-2 *Ontology – Part 2: Rules for developing ontologies in the Web Ontology Language (OWL)*
3. ISO 19150-3 *Ontology – Part 3: Semantic operators*
4. ISO 19150-4 *Ontology – Part 4: Service ontology*
5. ISO 19150-5 *Ontology – Part 5: Domain ontology registry*
6. ISO 19150-6 *Ontology – Part 6: Service ontology registry.*

Parts 1, 2, and 4 have been published. Part 6 is under development.

ISO 19150-1 sets the framework for semantic interoperability.

ISO 19150-2 defines the rules for converting UML static views of geographic information and application schemas to OWL ontologies.

ISO 19150-4 is modeled as an extension of the ISO 19101-1 *Reference model – Part 1: Fundamentals*. ISO 19150-4 makes use of service metadata (ISO 19115-1) and service definition (ISO 19119) whenever appropriate.

15.3.15 Ubiquitous Public Access (UPA)

The goal of the "ubiquitous public access to geographic information" (UPA-to-GI) concept can be described as making the user experience of any "smart" device intuitive to understand, along with being easy to use. The users are no longer passive consumers of geographic information, but rather become active participants in the entire lifecycle of geographic information, from collection via creation to dissimination.

Ubiquitous Public Access (ISO 19154)
ISO 19154 *Ubiquitous public access – Reference model* establishes a framework for a ubiquitous public access environment, in which ubiquitous public access to geographic information can be understood by three concepts: (1) ubiquity, (2) public access, and (3) ubiquitous public access [80].

Ubiquity, or omnipresence, is the property of being present everywhere. In geographic information, it is an environment to provide geographic information in every place at any time for any device.

Ubiquitous Public Access – Air Quality (ISO/TR 19167)
ISO/TR 19167 *The application of ubiquitous public access-to-geographic information for air quality information* is a technical report that aims to assist the dissemination of the ubiquitous public access context information model and to illustrate its application in the domain of atmospheric environments. Following the example of the city of Seoul the air pollutants PM (particulate matter), O_3, NO_2, CO, and SO_2 were monitored.

15.3.16 Preservation

Today, the vast majority of data is processed and stored digitally, including without doubt and in particular, spatial data. Unfortunately, the lifetime of digital storage media is much

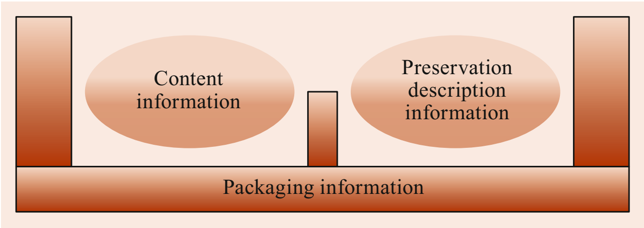


Fig. 15.42 Information package with content and preservation description information. (Adapted from [81])

shorter than the lifetime of analogue media, i.e. a few years, maybe up to 10 years, compared to hundreds of years in the case of paper, to name the most important example of analogue media. Another problem with the storage of digital media arises because, often, the data content is not self-explaining. Additional information is needed to fully access, understand, and use stored data. Thus, the usability of stored data heavily depends on the availability of complete and correct metadata. Obviously, digital data are not always stored in a way that adequately responds to media decay and the demand of long-term interpretability. Therefore, the era we live in is sometimes named the "digital dark age".

Several standardization organizations, including ISO, have started to address the challenge by developing standards that are to guarantee a long-term preservation of selected data and avoid uncontrolled loss. Typically, a classical analogue archive was the example. Starting from there, its role is projected to the specific aspects of digital environments. An archive in the conventional sense is a government organization, institution, or company that takes ownership of the archival material and ensures understandability, preservation of content, integrity, and authenticity under the stuartship of skilled experts, the archivists.

A digital archive operates differently in several ways. Responsibility regarding content is shifted to the submitter, while the archivists tend to become IT administrators. Their tasks include guaranteeing the exclusion of data loss and the management of digital rights in the case of data consumption.

In the field of geographic information, two major standards and one minor standard have been developed: ISO 14721 *Open archival information system (OAIS) – Reference model*, the 19165-1 *Preservation of digital data and metadata – Part 1: Fundamentals*, and ISO 19165-2 *Preservation of digital data and metadata – Part 2: Content specification for earth observation data and derived digital products*.

Open Archival Information System (OAIS) – Reference Model (ISO 14721)

ISO 14721:2012 *Space data and information transfer systems – Open archival information system (OAIS) – Reference model* (full title) was developed by the Consultative Committee for Space Data Systems (CCSDS). This group originated from working groups of ESA (Europe) and NASA (US), and has become a world-wide standardization body for space data and information systems.

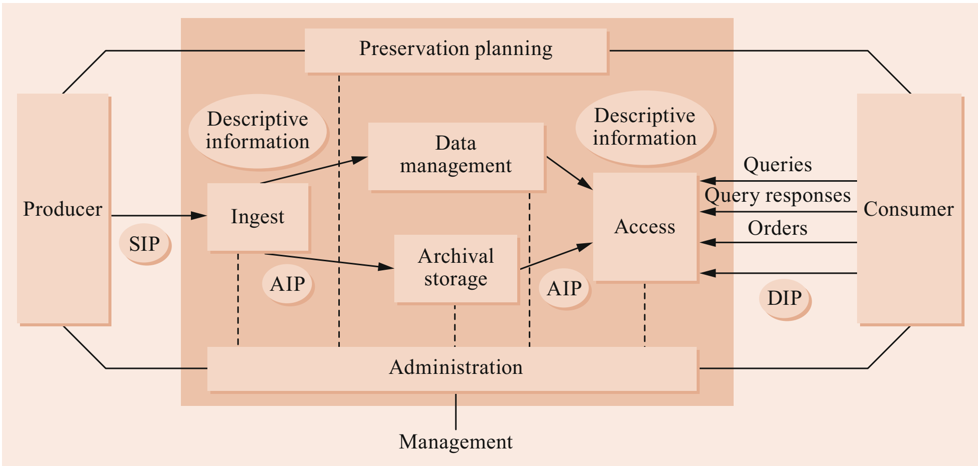
The ISO 14721 defines a number of structural elements that have been widely accepted by digital libraries. The information package (IP) is a logical container of two types of information called content information and preservation description information. Both are encapsulated in a package and identifiable by the packaging information (Fig. 15.42). The resulting package is viewed as being discoverable by the packaging's descriptive information [81].

The Open Archival Information System (OAIS) defines the data flow from the producer to the archive and the archive's interaction with the consumer. The producer sends a submission information package (SIP) to the archive, where the ingest functionality entity provides the services and functions to accept the SIP. Ingest functions include performing quality assurance and generating an archival information package (AIP). The access functional entity supports the consumer in finding requested information and generates the dissemination information package (DIP) for delivery of the data (Fig. 15.43).

The designated community is an identified group of potential consumers who should be able to understand a partic-

15

Fig. 15.43 OAIS functional entities. (Adapted from [81])



ular set of information. OAIS makes the content information in its AIP visible and available to its designated communities.

Preservation – Part 1: Fundamentals (ISO 19165-1)
The previously discussed ISO 14721 does not completely cover all the needs for digital data and metadata preservation for geospatial data in general. Therefore, the 19165-1:2018 *Preservation of digital data and metadata – Part 1: Fundamentals* addresses geospatial data, its data model structures, the multiplicity of data formats, and intellectual property rights. ISO 19165-1 is modeled as a specialization of ISO 19115-1 *Metadata – Part: Fundamentals*. The standard's development was initiated by national mapping and cadastral agencies and the Open Geospatial Consortium [82].

ISO 19165-1 adopts the concept of an information package defined in ISO 14721 and defines the elements of the IP as one of its central components. This package should be able to be shared with other organizations, including those outside the geospatial community.

ISO 19165-1 specializes the versions of the IPs, which are named geospatial submission information package (geo-SIP), geospatial archival information package (geo-AIP), and geospatial dissemination information package (geo-DIP). Their properties include lossless compression, cartographic series support, i.e., a manageable regional size, as well as a container for information regarding geometry (vector and raster), attributes, topology, metadata, quicklooks, and recommendations on how to symbolize the data [81].

The way geospatial data are disseminated on the Internet has forced a separation between data and metadata for practical reasons. The OAIS information package offers us a mechanism to use a "container" (information package) where data and metadata are kept together. ISO 19165-1 proposes to use the Open Packaging Conventions (OPC) defined in ISO/IEC 29500-2:2012 to build a geospatial information package. The OPC standard can be considered a modern version of the TAR format: it combines a ZIP compression of the parts composing the package (respecting a directory structure) with XML documents that describe the package content.

Preservation – Part 2: Earth Observation Data (ISO 19165-2)
The ISO 19165-2 *Preservation of digital data and metadata – Part 2: Content specifications for earth observation data and derived digital products* is an extension of the ISO 19165-1. It contains a more detailed specification of remote-sensing data, such as calibration data, full provenance (data history), and needs for the designated community.

The Earth Observation Preserved Dataset content consists of mission concept (MC), mission definition (MD), mission implementation (MI), mission operations (MO), and post-mission (PM).

15.3.17 Building Information Modeling

Building information modeling (BIM) is a planning method based on digital models of buildings. buildingSmart is a world-wide operating organization founded in 1995 and dedicated to the development of openBIM standards for civil infrastructure and building asset information. Their suite of standards has been adopted by ISO as ISO 16739:2013 *Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries*.

Obviously, the worlds of geographic information and building information modeling are rather similar. They both use vector and raster geometries, including reference systems, they apply object-oriented modeling and they vizualize their data. However, models and rules were developed independently over the last more than 20 years.

The standard ISO/TS 19166:2021 *BIM to GIS conceptual mapping (B2GM)* aims at investigating interoperability barriers with the goal of projects moving seamlessly from planning to civil infrastructure design. The work includes an analysis of the differences between the general feature model defined in ISO 19109 and the IFC metamodel, a comparison of the coordinate reference systems, a check of terminology, and a comparison of the vector geometries defined in ISO 19107 *Spatial schema* with those of ISO 10303-42:2014 *Industrial automation systems and integration – Product data representation and exchange – Part 42: Integrated generic resource: Geometric and topological representation*.

The standard is subdivided into three parts:

- BIM to GIS Perspective Definition (B2G PD): supports perspective information representation depending on the specific requirement such as the urban facility management. "Perspective" depends on the use-case. For example, to manage the urban facilities, the required data should be collected from the various data sources including BIM model and transformed to represent in userspecific perspective. B2G PD defines a Data View to extract the needed data and transform the information from the various data sources.
- BIM to GIS Element Mapping (B2G EM): supports the element mapping from BIM model to GIS model. As the BIM and GIS model schemas are different, B2G EM requires a mapping rule specifying how to transform from BIM model to GIS model element.
- BIM to GIS LOD Mapping (B2G LM): supports the LOD mapping from BIM model to GIS model. LOD (Levels Of Detail) in GIS is a deliberate choice of data included/excluded from a model to satisfy a certain use

cases including visualization. The relevant geometric and other information for the LODs required in the target GIS model need to be extracted/or queried from the BIM model. This can be defined by the LOD mapping rule set.

Further details can be found in Chap. 21.

15.3.18 Theme-Oriented Standards

The definition of catogories helps to better understand the multiplicity of ISO standards for geographic information together with their diverse numbers. The major categories discussed earlier in this chapter covered the reference model, the foundations of data modeling, and geometry. In contrast to those, some of the ISO 19100 standards have a relevance that is limited to a specific field of application. The following section presents standards for location-based services (ISO 19132–ISO 19134), the classification system of the Food and Agricultural Organization of the UN (FAO) (ISO 19144-1/2), property cadastre (ISO 19152), and addressing (ISO 19160-1/3/4).

Location-Based Services (LBS)

Location-based services are addressed in several standards of geographic information. In detail, these are the reference model (ISO 19132), tracking and navigation (ISO 19133), multimodal operation (ISO 19134), and three other standards regarding the movement along linear features.

LBS – Reference Model (ISO 19132)

ISO 19132 defines a reference model for location-based services (LBS) [83]. This reference models consists of a so-called conceptual framework, a frame of concepts, ideas, terms, definitions, and the interdependence between those. It also addresses the interfaces for data access while roaming and areas where further standards are required.

The role that ISO 19132 plays for LBSs is identical to the role that ISO 19101-1 *Reference model – Part 1: Fundamentals* plays for GIS.

ISO 19132 builds upon the viewpoints as defined in the Reference Model for Open Distributed Processing (RM-ODP; ISO/IEC 10746) but limits their application for an LBS system to the first three, the enterprise viewpoint, the information viewpoint, and the computational viewpoint. The other two, the engineering and technology viewpoints, play a subordinate role in this standard.

LBS – Tracking and Navigation (ISO 19133)

Tracking and navigation are the core applications of location-based mobile services. They are covered by ISO 19133 *Location-based services – Tracking and navigation* [84]. Tracking is the process of following and reporting the position of a vehicle in a network. In some cases, it may be the position of a handheld computer only. Routing is the finding of optimal routes between positions in a network. Route traversal is the execution of a route, usually through the use of instructions at each node in the path, and a start and stop instruction, at the first and last position of the route. The combination of routing, route transversal, and tracking is navigation. The optimal route is the one with minimal costs in terms of money, time, or other parameters, such as pleasure along scenic routes.

Tracking

The tracked positions are defined by coordinates or other types of positional descriptions. A tracking service (class TK_TrackingService) delivers the positions, either one by one or as a sequential list.

The positions may be one of the types coordinate, place name, feature, linear reference, network, address, or phone.

A trigger defines the moment or the location for delivering positional information. Triggers are generally of two types: triggered by an event or triggered by the passage of time.

A transitional trigger delivers a new position dependent on the movement of the vehicle being tracked. Usually, events take place after the completion of a distance, or after a change of direction. The periodic trigger is used to control location sequences by setting temporal limits on how far apart in time tracking samples are taken.

The tracking location metadata include the mobile subscriber and the quality of positions. The mobile subscriber is the item being tracked such as a car with a navigation system.

Navigation

Navigational computation is based on an underlying network. The network comprises the elements:

- The nodes represent important points in the network, such as intersections.
- The links represent uninterrupted paths between nodes with an orientation that indicates which direction the link is to be traversed.
- The turns associate a node to an entry link and exit link to a node.
- The stops consist of either nodes or positions on links within the network. The start and the end of the route are a so-called type of stop.

According to ISO 19133, a network has two different topologies. Its geometric topology refers only to nodes and edges. The second topology is the graph of links, junctions, and turns that is shown in Fig. 15.44. Although the links, junctions, and turns have the same underlying geometry, they have their own connectivity based on usable *vehicle* routes.

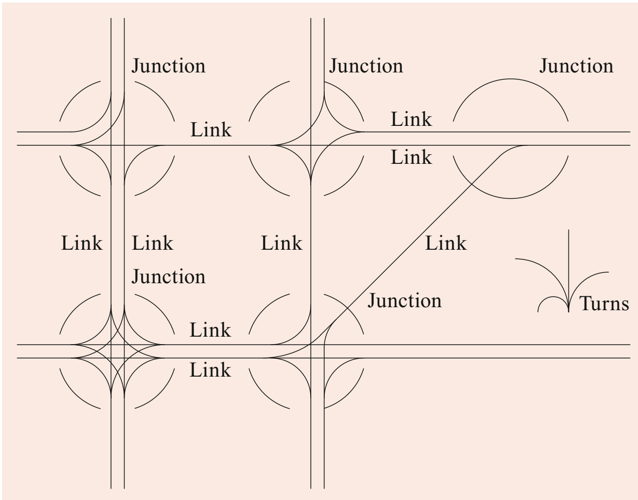


Fig. 15.44 Topology of a road network with links (edges), junctions (nodes), and turns

Table 15.6 Attributes of a route request

Route request type	Basic, predictive, etc.
Vehicle	Type, such as car or truck
Way point list	Start point, end point, other stopping points
Avoid list	Links in the network to be avoided
Departure time	Planned period of the beginning of navigation
Arrival time	Planned period of the end of navigation
Cost function	Type, default is minimum distance
Preferences	Specific user demands such as most scenic route
Advisories	Text to be displayed during navigation
Refresh interval	Maximum time before a recalculation of the route

If a link comes to a crossroads, and U-turns are allowed, then there are up to four turns that exit that link.

The description of a network includes constraints, such as vehicle constraints, temporal constraints, and lane constraints.

A navigation service delivers the optimal route between two positions within a network and guides the vehicle along that route. A navigation service requests a route from the navigation system and receives a proposed route based on the given parameters. Table 15.6 presents the most important attributes for a route request.

The cost function calculates the optimal route based on minimum costs. ISO 19133 recommends the algorithms of *Dijkstra* [85] and *Bellman–Ford* [86, 87]. Table 15.7 summarizes the typical variables used to control a cost function for car navigation.

LBS – Multimodal Operation (ISO 19134)

ISO 19134 *Location-based services – Multimodal routing and navigation* extends ISO 19133 *Location-based services – Tracking and navigation* in that is adds the case of a multimodal network [88]. This type of network is the typical setup of urban public transport.

Table 15.7 Cost function variables for route calculation

Distance
Time
Stopping time (traffic lights)
Speed
Speed limits
Slope (affects mostly freight vehicles)
Link capacity (capacity of an intersection)
Link volume (current amount of traffic at an intersection)
Stopping time (expected stopping time at turns)
Conditions (weather)
Tolls

A multimodal network is a network that is composed of edges that belong to more than one traveling mode, e.g., bus and light rail. If a network consists of one means of traveling, it only has single-mode junctions.

Movement Along Linear Features

This section addresses the standards *Linear Referencing* (ISO 19148), *Schema for Moving Features* (ISO 19141), and *Transfer Nodes* (ISO 19147).

Linear Referencing (ISO 19148)

ISO 19148 *Linear referencing* defines a model for relating attributes and locations to linear geographic features, such as roads and pipelines, or said in a more general way, transportation and utilities [89]. This standard is a specialization of ISO 19133 *Location-based services – Tracking and navigation*. It is also based on the family of standards that was developed by ISO/TC 204 *Intelligent transport systems*.

Worldwide, there are numerous methods used for relating events such as road accidents and utility leakages to a linear feature, and, thus, there is no single best method. Therefore, ISO 19133 can only provide a generalized model.

The model has a number of characteristics:

- The features do not need to be a linear geometry, but must allow for measurements in a one-dimensional, linear sense. This refers to the abstraction of a road network as used in navigation systems and to the fact that a road is never one-dimensional, as it has a width different from zero.
- The feature may have no geometry but only a reference to one. Maintenance activities such as snow plowing may be an example of this.

Point positions may be referenced (the standard says "projected") to the linear feature.

A linear element may be a feature as defined in this standard, a curve, which is a geometry primitive defined in ISO

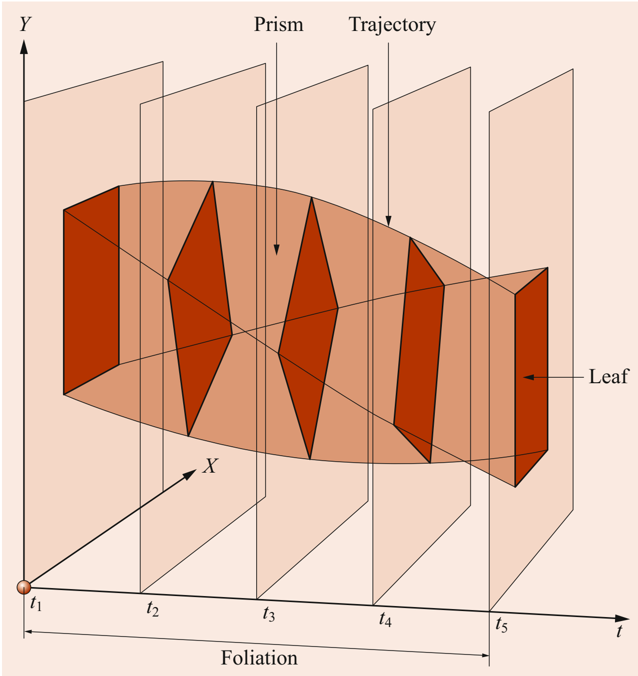


Fig. 15.45 Graphical explanation of the terms foliation, prism, trajectory, and leaf. (Adapted from [90])

19107, or a directed edge, which is a topological primitive defined in ISO 19107.

Schema for Moving Features (ISO 19141)

ISO 19141 *Schema for moving features* defines a model for the description of features in motion. Such a feature may be a car, an airplane, a ship, or something similar. While moving along the trajectory of the feature, the feature's orientation/ rotation and the clearance gauge, a measurement instrument to determine the clearance, are of interest. ISO 19141 addresses those elements.

Fig. 15.45 illustrates how the concepts of foliation, prism, trajectory, and leaf relate to one another. In this illustration, a 2-D rectangle moves and rotates, and represents such a moving feature. Each representation of the rectangle at a given time is a leaf. The path traced by each corner point of the rectangle (and by each of its other points) is a trajectory. The set of points contained in all of the leaves, and in all of the trajectories, forms a prism. The set of leaves also forms a foliation.

Transfer Nodes (ISO 19147)

ISO 19147 *Transfer Nodes* supports planning authorities to model multimodal transport systems. A transfer node is a location that establishes a link between two of more means of transportation. Typically, this is a station that links long-distance trains with commuter trains or the road with the railroad network. ISO 19147 defines transfer nodes and their attributes, a feature catalogue for transfer nodes, and how transfer nodes can be linked to real-world locations [91].

Landcover Classification Systems

Classification Systems – Part 1: Classification System Structure (ISO 19144-1)

Classification is the grouping of things or terms according to features that they have in common [92] and is a basic step in most sciences. In geography, the classification of regions based on thematic phenomena has always been an important research topic. The Troll–Paffen map of the classification of the Earth's climate zones is a prominent example [93].

Classification in geographic information sciences faces the challenge of defining and applying classification schemas in the Internet era. The ISO 19144-1/2 group of standards entitled classification systems sets basic rules for structuring those schemas and defines a so-called land cover metalanguage that shall be used for writing such a schema.

ISO 19144-1 does not define a single classification system as the *standard one* but rather sets rules for creating such a system. ISO 19135-1 *Procedures for item registration* provides the rules for those systems to allow for a well-organized and easy-to-maintain system.

The classification system subdivides an area into small units, each of which carry an identification code. Using the terminology of ISO 19123 *Schema for coverage and functions* (*coverage* in short) this partitioning is called a *discrete coverage*. Therefore, most of the classification methods result in discrete coverages.

A number of classification characteristics are worth mentioning.

A geographic feature according to ISO 19101-1 *Reference model – Part 1: Fundamentals* is different from a classified object according to ISO 19144-1 *Classification Systems – Part 1: Classification system structure* in that a geographic feature is atomic, while a classified object is part of a whole and, thus, necessarily related to the others by the classifiers that decompose the whole. The standard gives the example of classifying the Earth's surface into land and water. The total of land and water is the complete surface.

As a classification is an abstract representation of real-world phenomena, a classification system shall be:

- Scale independent, meaning that the classes at all levels of the system shall be applicable at any scale or level of detail.
- Source independent, implying that it is independent of the means used to collect information.

Classification systems come in two basic forms: hierarchical and nonhierarchical.

The requirements are the following:

- For nonhierarchical classification systems
  - Classifiers shall be defined so that all classes are mutually exclusive.
- For hierarchical classification systems
  - Classifiers shall be defined so that all classes at a specific level of the hierarchy are mutually exclusive.
  - Criteria used to define a classifier at one level of a hierarchical classification shall not be repeated at another level (e.g., criteria used to define a classifier at a lower level shall not be duplicated at a higher level of the hierarchy).

Classification Systems – Part 2: Land Cover Metalanguage (LCML) (ISO 19144-2)

Land cover is the (bio)physical cover on the Earth's surface. This definition is provided by ISO 19144-2 *Classification systems – Part 2: Land cover metalanguage (LCML)*. Land cover is a geographically explicit feature that other disciplines may use as a geographical reference, e.g., for land use, and climatic or ecological studies.

Many land-cover classification systems (LCCSs) are in use today and will remain so in the future because they serve specific needs, and investments have been made to exploit their data. Examples of well-known international LCCSs are the UN/Food and Agriculture Organization (FAO) LCCS, Coordination of Information on the Environment (CORINE, European Union), Africover (United Nations), Anderson (United States Geological Survey (USGS)), and Global Map (Japan), as well as many national developments. ISO19144-2 provides a metalanguage expressed as a UML model that allows different land cover classification systems to be described. This metalanguage allows the description of land cover features based on physiognomy, i.e., the general appearance of an object or terrain, without reference to its underlying or scientific characteristics. Those features may be part of different land cover legends (nomenclature), thus providing a framework to compare different systems.

The aim of ISO 19144-2 is to enable the ability to compare information from existing classification systems in a meaningful way without replacing them. The aim is to complement the development of future classification systems that may offer more reliable collection methods for particular national or regional purposes by allowing them to be described in a consistent manner. It provides a common reference structure for comparison and integration of data for any generic land cover classification system but is not intended to replace those classification systems.

Property Cadastre: Land Administration (ISO 19152)

ISO 19152 *Land administration domain model (LADM)* is intended to help build modern property cadastre systems, in particular in those countries that do not have a century (or more) of cadastral tradition. ISO 19152 is based on the experiences of the advanced countries combined with the data models of the ISO/TC 211 family of standards [94].

ISO 19152 defines a reference LADM covering all basic information-related components of land administration, including those over water as well as land, and elements above and below the surface of the Earth. The standard provides an abstract, conceptual schema with five basic packages related to:

1. Parties (*people and organizations*)
2. Rights, responsibilities, and restrictions (*ownership rights*)
3. Spatial units (*parcels, buildings, and networks*)
4. Spatial sources (*surveying)*
5. Spatial descriptions (*geometry and topology*).

This standard also defines a terminology for land administration, based on various national and international systems, which is as simple as possible in order to be useful in practice and includes the basis for national and regional profiles.

Property cadastre and land administration are usually closely linked to national legislation. This standard is not intended to interfere with any national land administration laws. Furthermore, the construction of external databases with party data, address data, valuation data, land use data, land cover data, physical network data, and taxation data is out of its scope. However, ISO 19152 provides blueprint stereotype classes for these datasets, which indicate which dataset elements the land administration domain model expects from these external sources, if available.

Fig. 15.46 depicts a comprehensive picture of the LADM. The LADM is also addressed in Chap. 20.

Addressing

The series of the ISO 19160-1/3/4 standards defines a framework for activities involving addresses. Addresses are a way to unambiguously determine an object for purposes of identification and location. The arrangement of addresses varies from country to country. While in Europe and America the road network forms the primary level to define a location, other systems are in place. For instance, in Japan, administrative areas serve as the primary level. Other parts of the world are still in an early stage of developing nationwide postal systems.

The purpose of address systems includes postal delivery, emergency response, customer contacts, land administration, as well as utility planning and maintenance. Consequently, a heterogeneous group of stakeholders is using this standard for various purposes: local and national governments and postal operators, electronic business, election commissions, and many others. The prime stakeholder was the Universal Postal Union (UPU), which was founded under the name "Weltpostverein" in 1874, proposed by the German Heinrich von Stephan (French name: Union postale universelle). The

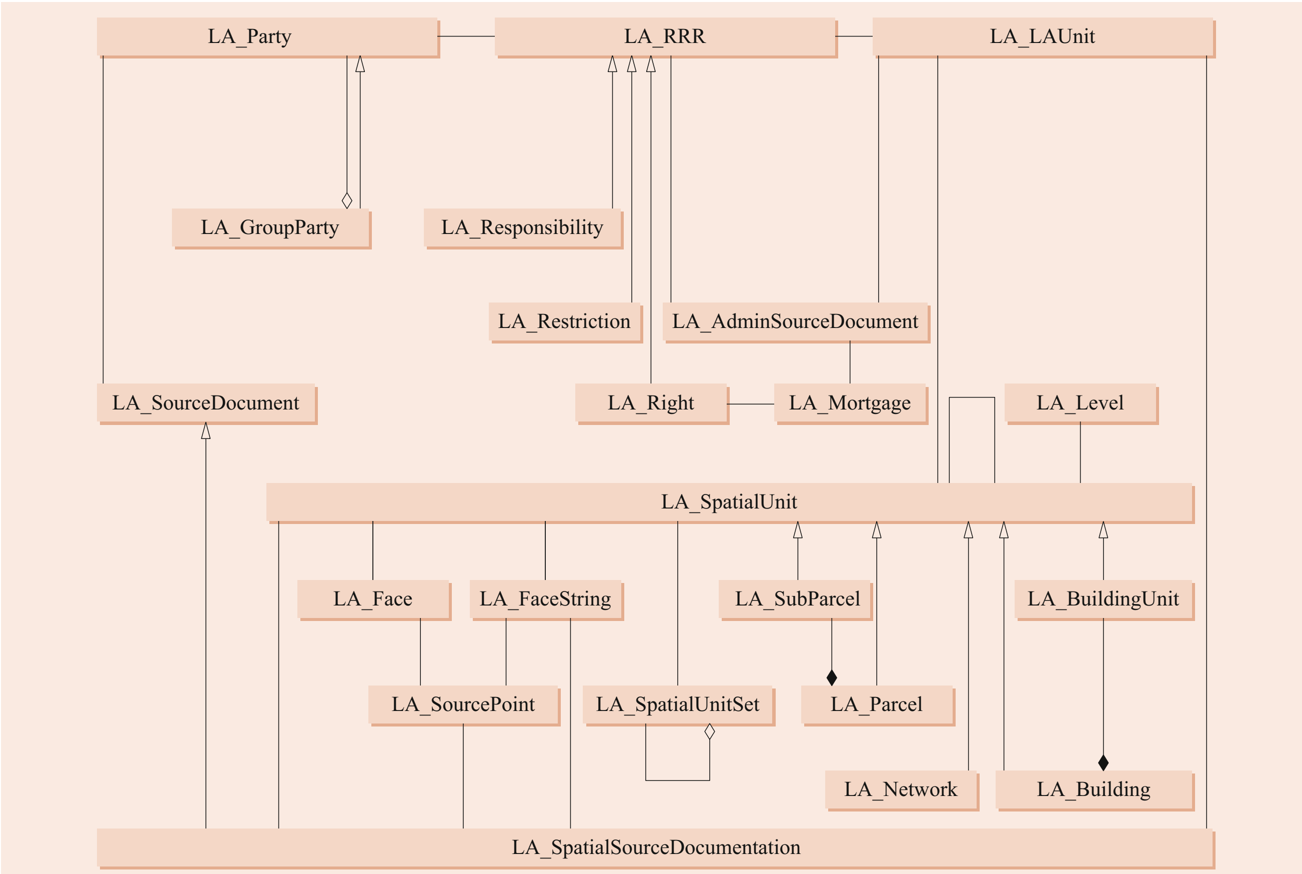


Fig. 15.46 Land administration domain model. LA_AdminSourceDocument – source document such as a contract, LA_Building – a description of the legal, recorded, or informal space, not the physical entity, LA_BuildingUnit – a part of a building according to LABuilding, LA_Face – two-dimensional topological primitive (surface), also used to build 3-D geometries, LA_FaceString – line information (geometry), LA_GroupParty – group of parties, e.g., an association, LA_LAUnit – generalized unit that is, e.g., used to associate right, LA_Level – content level such as urban, rural, etc., LA_Mortgage – details of the mortgage, LA_Network – information about networks such as electricity, LA_Parcel – spatial unit with a legal ownership right, LA_Party – involved player, such as a notary, LA_Responsibility – responsibility such as maintenance of a waterway, LA_Restriction – restriction such as servitude, LA_Right – rights such as ownership, LA_RRR – rights, restrictions, responsibilities, LA_SpatialUnit – a superclass covering parcels, buildings, etc., LA_SpatialUnitSet – set of spatial units such as a municipality, LA_SubParcel – subunit of a parcel according to LA Parcel, LA_SourcePoint – point information (geometry), LA_SpatialSourceDocument – documentation of land administration surveys (Adapted from [94])

goal of the addressing standards is facilitating interoperability between existing and future address specifications.

The address standards consist of a set of five individual standards:

- ISO 19160-1 *Addressing – Conceptual model*
- ISO 19160-2 *Addressing – Good practices for address assignment schemes*
- ISO 19160-3 *Addressing – Quality management for address data*
- ISO 19160-4 *Addressing – International postal address components and template language*
- ISO 19160-5 *Addressing – Address rendering for purposes other than mail*
- ISO 19160-6 *Addressing – Digital interchange models*.

In the early stage of development of the ISO 19100 series of standards, a similar standard was published. This is ISO 19112:2019 *Spatial referencing by geographic identifiers*. Though the requirements for the addressing standards are sufficiently different from the scope of ISO 19112, this older standard may be expressed as a profile of the addressing standards in the future.

At the time of publication of this Handbook, parts 1, 3, and 4 have been published as international standards.

Part 1 Conceptual model: this defines six conformance classes that are used to evaluate and control conformance of an existing or future addressing system to these addressing standards [95]. The conformance classes are: core, lifecycle, provenance, locale, full conformance, and address profile documentation.

Part 3 Quality management for address data: this sets the addressing standards as a specialization of the ISO 19157 *Data quality*. Consequently, the addressing standards simply define a number of additional quality elements. These elements are grouped according to the ISO 19157 in completeness, logical consistency, positional accuracy, temporal quality, and thematic accuracy.

Part 4 International postal address components and template language: today, the vast majority of postal items are addressed in a way that can be processed automatically, allowing economic processing and minimizing the risk of misinterpretation. Those processes are based on address databases. Part 4 provides a language to express the creation of a specific address layout from the data stored in an address database. This process is also called rendering of an address. This standard is intended for designers and developers of computer systems that process address data.

15.3.19 Future of Geomatics Standardization

Directions and Activities

The work of ISO/TC 211 *Geographic information/Geomatics* has completed its first lifecycle. All base standards have been published and, consequently, moved to practical implementations with responsibilities devolving to industry and national or supranational government institutions. The earlier publications are presently under their first revision.

Trends

Though the future direction of ISO/TC 211 is not known in detail, the prevailing trends are clear. The ISO/TC 211-member countries are committed to continue adoption of the ISO 19100 standards. Many of the project teams are working on mostly minor issues that have come up after implementing the standards in comprehensive work environments. Examples of such work are the definition of XML interfaces for metadata and registries, or the merging of formerly three quality standards to a single future document. ISO 19100 standards model geographic information as a special case of information technology (IT). In the early stage of the Geographic Information standards many GI-specific tools needed a separate development, e.g. the web map server. In the meantime, IT has significantly advanced. Therefore, some GI-specific standards will be replaced by general IT technology.

There will be a strong case for using profiles and application schemas of the ISO 19100 family to build well-tailored environments for specific fields of applications. One example is the field of property cadastre.

ISO/TC 211 will also be shifting its focus to neighboring areas of work, where the industry has expressed a need for standardization. The most prominent example of such a new area is location-based mobile services.

Strength

ISO/TC 211 has become the worldwide-accepted umbrella group for standardization of geomatics. It has succeeded in engaging the national bodies and kindred organizations in the development of a suite of international standards. ISO/TC 211 has evolved as a competent group that is able to represent geomatics, fairly independently of national and industrial influences and with a sufficiently clear border to other fields of ISO.

Weakness

In geomatics, many private standards are already in place, with some of them being de facto standards. The newer ISO standards cannot replace them on a short-term basis, and it might, therefore, take a long period until the ISO standards are fully implemented.

Opportunities

ISO/TC 211 has formulated the abstract basis for the whole world of geomatics. Through its neutral position, it has become the coordinator of many activities in the geomatics business. Productive collaborations among liaison and national members have been or will be initiated. The first cooperative agreement was signed with the OGC, followed by the IHO. Other organizations such as the Defence Geospatial Information Working Group (DGIWG) have followed. Close cooperation with national associations is also anticipated.

The registry (ISO 19135-1) opens a flexible interface from the ISO 19100 standards to a wide range of worldwide applications. The registry will not only allow keeping track of ISO-compliant standards but will also be a continuously growing source of geomatics implementations beyond the scope of ISO/TC 211.

The ISO 19100 standards were chosen as the reference for a number of emerging technologies.

The Global Spatial Data Infrastructure (GSDI) is promoting a globally coordinated approach to geomatics.

The Infrastructure for Spatial Information in Europe (INSPIRE) sets the framework for a geomatics infrastructure in Europe. The United Nations (UN) are promoting application of the ISO standards among their members.

ISO/TC 211 has established an advisory group on strategy to address future directions and report to the chairman.

Promotional Activities

In the past, there were concerns about the ISO 19100 standards not being well known and, as a consequence, not widely used. ISO/TC 211 has, therefore, started an outreach campaign to create awareness of its standards.

The creation of awareness will support application-oriented activities, including implementation of transfer standards by vendors, implementation of data cataloguing standards by data producers, and implementation of metadata standards by vendors and general users. The development of awareness will include education and training, as well as the creation of user communities. ISO/TC 211 has established an advisory group on outreach to follow up the promotional activities.

Political Background

Many countries are understanding the tremendous value of GIS and their data for governing the Commonwealth. In the USA, homeland security takes a strong interest in building a geographic data infrastructure to improve emergency management, including the fight against terrorism and better disaster management. The USA also supports the development of a spatial data infrastructure (SDI), because GIS is believed to offer an effective weapon against global terrorism.

The development of the European SDI is governed by the long tradition and high quality of the property cadastre and topographic portrayal on this continent. Another strong component is led by the need for environmental protection. The INSPIRE initiative provides an integrated approach for all countries of the European Union.

Implementation Strategies

ISO/TC 211 has about 70 member countries and 35 external liaison organizations. While participating in the work of ISO/TC 211, the members committed themselves to adopting the resulting standards in their home environment. However, the implementations will vary from country to country, from organization to organization, and in some cases, even from province to province.

Most implementations will not require the complete functionality of the ISO 19100 series. Instead, only a subset of the abstract standards will be needed. The implementations remain compatible as far as their application schemas overlap. The usage of ISO 19100 guarantees a common underlying abstract model that enables compatibility if the applications schemas are the same.

The OGC plays an important role at the implementation level. In theory, ISO/TC 211 develops the abstract standards, and the OGC develops the implementation standards. In practice, the borderline is not drawn that clearly.

The partitioning of the work between ISO/TC 211 and the OGC has turned out to be fruitful for the geomatics community. The development of an ISO standard may take up to 4 years. This is acceptable for long-term abstract standards that enjoy a consensus among the large international community, but that time is far too long for implemented computer software. Programmers would just ignore the official standards. Implementation specifications became the domain of the OGC. An example is GML version 3.x (GML 3.x). GML 3.x is a fairly complete implementation of ISO 19107 *Spatial schema*. It was questioned whether GML 3.0 needed to become an ISO standard too. However, because many nations only recognize ISO standards, GML became ISO 19136 in 2007.

Many member countries and liaison organizations of ISO/TC 211 are adopting the implementation of the ISO 19100 standards from OGC.

15.4 Liaison Members of ISO/TC 211

15.4.1 Internal Liaison Members of ISO/TC 211

ISO/TC 211 *Geographic information/Geomatics* was founded in 1994 to standardize geographic information. However, various aspects of geographic information had been published as ISO, IEC, or ITU standards before the ISO/TC 211 was founded, and other relevant standards are being developed while the work of ISO/TC 211 is ongoing. Formal relations with the responsible committees, known as *internal liaisons*, have been established in order to align the existing standards and ongoing activities with the work of ISO/TC 211.

The number of internal liaisons varies according to the needs of the standardization work. The following list has been assembled from *all* technical committees and subcommittees of ISO and IEC, as well as all study groups of ITU.

In many cases, only a very specific working group or subcommittee of a technical committee will develop an internal liaison relationship with the ISO/TC 211. An example is ISO/TC 20 *Aircraft and space vehicles*. Only subcommittee 13 (ISO/TC 20/SC 13) entitled *Space data and information transfer systems* has an internal liaison with ISO/TC 211. The working groups and the subcommittees relevant to geographic information are titled accordingly and placed as the first paragraph of each section. In situations where the names of the working groups and subcommittees do not clearly define the subjects that they cover, a list of topics is provided for clarification.

ISO

- *ISO/TC 20: Aircraft and space vehicles.* Subcommittee relevant to geographic information
  - Space data and information transfer systems (SC 13)
  - Space systems and operations (SC 14)
  - Unmanned aircraft systems (SC 16)
- *ISO/TC 23: Tractors and machinery for agriculture and forestry*. Subcommittee relevant to geographic information
  - Agricultural electronics (SC 19)
- *ISO/TC 46: Information and documentation*

- *ISO/TC 59: Buildings and civil engineering works.* Subcommittee relevant to geographic information
  - Organization of information about construction works (SC 13)
- *ISO/TC 69: Application of statistical methods*
- *ISO/TC 127/SC 3: Machine characteristics, electrical and electronic systems, operation and maintenance*
- *ISO/TC 154: Processes, data elements and documents in commerce, industry and administration*
- *ISO/TC 172: Optics and photonics.* Subcommittees relevant to geographic information
  - Geodetic and surveying instruments (SC 6)
- *ISO/TC 184: Automation systems and integration.* Subcommittee relevant to geographic information
  - Industrial data (SC 4)
- *ISO/TC 190: Soil quality.* Subcommittee relevant to geographic information
  - Evaluation of criteria, terminology and codification (SC 1)
- *ISO/TC 204: Intelligent transport systems*
- *ISO/TC 207: Environmental management.* Subcommittee relevant to geographic information
  - Greenhouse gas management and related activities (SC 7)
- *ISO/TC 241: Road traffic safety management system*
- *ISO/TC 268: Sustainable cities and communities*
- *ISO/TC 292: Security and resilience*
- *ISO/TC 307: Blockchain and distributed ledger technologies*

ISO/IEC JTC1

- *ISO/IEC JTC 1/WG 9: Big Data*
- *ISO/IEC JTC 1/WG 11: Smart Cities*
- *ISO/IEC JTC 1/SC 2: Coded character sets*
- *ISO/IEC JTC 1/SC 24: Computer graphics, image processing and environmental data representation*
- *ISO/IEC JTC 1/SC 29: Coding of still pictures (WG1)*
- *ISO/IEC JTC 1/SC 32: Data management and Interchange*
- *ISO/IEC JTC 1/SC 36: Information technology for learning, education and training*
- *ISO/IEC JTC 1/SC 41: Internet of Things and related technologies*
- *ISO/IEC JTC 1/SC 42: Artificial Intelligence*

CEN/TC 287

Standardization of geographic information in Europe started in 1991. Led by the Association Française de Normalisation (AFNOR), the national standards body of France, the European standardization organization CEN created its Technical Committee 287 Geographic information.

The cooperation between ISO and CEN is arranged in the *Vienna agreement* with its present version of 2001.

The agreement underlines that international standards take precedence over national standards. In an ideal case, ISO members align their own standardization with ISO, so that completed international standards can be simultaneously adopted as national standards. If an international standard is simultaneously approved as a CEN standard (Europe), it automatically becomes a national standard in all CEN member countries.

On the other hand, the agreement recognizes that the European market may have particular needs, for example:

- Standards for which there is no international need currently recognized
- Standards that are required in the European Union but have a lower priority at the international level [96].

Table 15.8 lists the profile of the ISO geomatics standards that have been approved by CEN/TC 287 since its revitalization. This profile is focused on the standards required for building the INSPIRE project.

15.4.2 External Liaison Organizations to ISO/TC 211

Portrait of All (Status November 2021)

The work of standardization committees relies heavily on the expertise of organizations outside the usual standardization business. These organizations, known as *external liaison members*, are assigned to the technical committees of ISO or IEC according to the needs of a specific standards development project.

Even though the external liaison members of ISO/TC 211 are all international representatives of orgainzations related to geographic information, they are different types of organizations with different organizational structures. They might be categorized in the following way.

International organizations, representing:

- Government agencies, e.g., IHO (hydrography) and the World Meteorological Organization (WMO) (weather).
- Scientific subjects, e.g., the International Association of Geodesy (IAG) (geodesy) and the International Federation of Surveyors (FIG) (surveying).
- Regional interests, e.g., the Joint Research Centre (JRC) (Europe) and the Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP) (Asia, Pacific).
- Some tasks of the United Nations, e.g., the United Nations Group of Experts on Geographical Names (UNGEGN) (geographic names).

The following list is a complete summary of all current external liaison members of ISO/TC 211. Internet addresses are

Table 15.8 Standards and other deliverables of CEN/TC 287 (EN = CEN-standard)

EN ISO 6709:2009	*Standard representation of geographic point location by coordinates* (ISO 6709:2008)
CEN/TR 15449-1:2012	*Spatial data infrastructure – Part 1: Reference model*
CEN/TR 15449-2:2012	*Spatial data infrastructure – Part 2: Besst practices*
CEN/TR 15449-3:2012	*Spatial data infrastructure – Part 3: Data centric view*
CEN/TR 15449-4:2014	*Spatial data infrastructure – Part 4: Service centric view*
CEN/TR 15449-5:2015	*Spatial data infrastructure – Part 5: Validation and testing*
EN ISO 19101-1:2014	*Reference model – Part 1: Fundamentals* (ISO 19101-1:2014)
EN ISO 19105:2005	*Conformance and testing* (ISO 19105:2000)
EN ISO 19106:2006	*Profiles* (ISO 19106:2004)
EN ISO 19107:2005	*Spatial schema* (ISO 19107:2003)
EN ISO 19108:2005	*Temporal schema* (ISO 19108:2002)
EN ISO 19108:2005/AC:2008	*Temporal schema* (ISO 19108:2002/Cor 1:2006)
EN ISO 19109:2015	*Rules for application schema* (ISO 19109:2015)
EN ISO 19110:2016	*Methodology for feature cataloguing* (ISO 19110:2016)
EN ISO 19111:2007	*Spatial referencing by coordinates* (ISO 19111:2007)
EN ISO 19111-2:2012	*Spatial referencing by coordinates – Part 2: Extension for parametric values* (ISO 19111-2:2009)
EN ISO 19112:2005	*Spatial referencing by geographic identifiers* (ISO 19112:2003)
EN ISO 19115-1:2014	*Metadata – Part 1: Fundamentals* (ISO 19115-1:2014)
EN ISO 19115-1:2014/A1:2018	*Metadata – Part 1: Fundamentals – Amendment 1* (ISO 19115-1:2014/Amd 1:2018)
EN ISO 19115-2:2010	*Metadata – Part 2: Extensions for imagery and gridded data* (ISO 19115-2:2009)
EN ISO 19116:2006	*Positioning services* (ISO 19116:2004)
EN ISO 19117:2014	*Portrayal* (ISO 19117:2012)
EN ISO 19118:2011	*Encoding* (ISO 19118:2011)
EN ISO 19119:2016	*Services* (ISO 19119:2016)
EN ISO 19123:2007	*Schema for coverage geometry and functions* (ISO 19123:2005)
EN ISO 19125-1:2006	*Simple feature access – Part 1: Common architecture* (ISO 19125-1:2004)
EN ISO 19125-2:2006	*Simple feature access – Part 2: SQL option* (ISO 19125-2:2004)
EN ISO 19126:2009	*Feature concept dictionaries and registers* (ISO 19126:2009)
EN ISO 19128:2008	*Web Map Server interface* (ISO 19128:2005)
EN ISO 19131:2008	*Data product specifications* (ISO 19131:2007)
EN ISO 19131:2008/A1:2011	*Data product specifications – Amendment 1: Requirements relating to the inclusion of an application schema and feature catalogue and the treatment of coverages in an application schema* (ISO 19131:2007/Amd 1:2011)
EN ISO 19132:2008	*Location-based services – Reference model* (ISO 19132:2007)
EN ISO 19133:2007	*Location-based services – Tracking and navigation* (ISO 19133:2005)
EN ISO 19134:2008	*Location-based services – Multimodal routing and navigation* (ISO 19134:2007)
EN ISO 19135-1:2015	*Procedures for item registration – Part 1: Fundamentals* (ISO 19135-1:2015)
EN ISO 19136:2009	*Geography Markup Language (GML)* (ISO 19136:2007)
EN ISO 19137:2008	*Core profile of the spatial schema* (ISO 19137:2007)
CEN ISO/TS 19139:2009	*Metadata – XML schema implementation* (ISO/TS 19139:2007)
EN ISO 19141:2009	*Schema for moving features* (ISO 19141:2008)
EN ISO 19142:2010	*Web Feature Service* (ISO 19142:2010)
EN ISO 19143:2012	*Filter encoding* (ISO 19143:2010)
EN ISO 19144-1:2012	*Classification systems – Part 1: Classification system structure* (ISO 19144-1:2009)
EN ISO 19144-1:2012/AC:2012	*Classification systems – Part 1: Classification system structure – Technical Corrigendum 1* (ISO 19144-1:2009/Cor 1:2012)
EN ISO 19146:2018	*Cross-domain vocabularies* (ISO 19146:2018)
EN ISO 19148:2012	*Linear referencing* (ISO 19146:2021)
EN ISO 19152:2012	*Land Administration Domain Model (LADM)* (ISO 19152:2012)
EN ISO 19156:2013	*Observations and measurements* (ISO 19156:2011)
EN ISO 19157:2013	*Data quality* (ISO 19157:2013)
EN ISO 19157:2013/A1:2018	*Data quality – Amendment 1: Describing data quality using coverages* (ISO 19157:2013/Amd 1:2018)

Source: https://standards.cen.eu/dyn/www/f?p=CENWEB:105::RESET:::: (7.9.2018) provided for further information. As an introduction, a listing of all external liaison members is given in brief:

- CalConnect The Calender and Scheduling Consortium
- CEOS/WGISS Committee on Earth Observation Satellites/Working Group on Information Systems and Services
- DGIWG Defence Geospatial Information Working Group
- Energistics
- ESA European Space Agency
- EUROGI European Umbrella Organisation for Geographic Information
- EuroSDR European Spatial Data Research
- FAO/UN Food and Agriculture Organization of the United Nations
- FIG International Federation of Surveyors
- GRSS IEEE Geoscience and Remote Sensing Society
- IAG International Association of Geodesy
- ICA International Cartographic Association
- ICAO International Civil Aviation Organization
- IHO International Hydrographic Organization
- IOGP International Association of Oil and Gas Producers
- ISPRS International Society for Photogrammetry and Remote Sensing
- JRC European Commission Joint Research Centre
- OASIS Organization for the Advancement of Structured Information Standards
- OGC Open Geospatial Consortium, Incorporated
- OMG Object Management Group
- PAIGH Panamerican Institute of Geography and History
- SBS Small Business Standards
- SCAR Scientific Committee on Antarctic Research
- UN ECA United Nations Economic Commission for Africa
- UNECE United Nations Economical Commission for Europe, Statistical Division
- UNGEGN United Nations Group of Experts on Geographical Names
- UN-GGIM United Nations Committee of Experts on Global Geospatial Information Management
- UN-GGIM-AP United Nations Committee of Experts on Global Geospatial Information Management for Asia and the Pacific
- UPU Universal Postal Union
- WMO World Meteorological Organization
- World Bank.

CalConnect

The name CalConnect stands for the Calendaring and Scheduling Consortium. CalConnect focuses on the exchange of calendaring and scheduling information between dissimilar programs and platforms. Work started in the late 1990s under the auspices of the IETF (Internet Engineering Task Force). The consortium began its work in 2004:

- Objectives:
  - Promote calendaring and scheduling and support cooperation between members;
  - Conduct interoperability tests;
  - Develop calendaring and scheduling specifications;
- Members: about 30, mostly calendar software, IT, telecommunication, logistics
- Chair: US
- Homepage: https://www.calconnect.org/.

CEOS Committee on Earth Observation Satellites/ Working Group on Information Systems and Services (CEOS/WGISS)

CEOS was established in 1984 in response to a recommendation from the G7 Economic Summit. Since then, the circumstances surrounding the collection and use of space-based Earth observations have changed: The number of Earth-observing satellites has vastly increased. Onboard instruments are more complex and are capable of collecting new types of data in ever-growing volumes. The user community has expanded and become more diverse:

- Objectives:
  - Exchanging technical information
  - Increasing data usefulness and cost effectiveness
  - Presenting plans for emerging technologies and programs
  - Addressing current developments and future opportunities
  - Raising awareness in and having discussions with user communities.

WGISS is one of the five working groups of CEOS (the other four being the Working Groups on calibration and validation, on capacity building and data democracy, on climate, and on disasters). WGISS aims to stimulate, coordinate, and monitor EO initiatives, thereby enabling users at global, regional, and local levels to exploit more effectively and benefit from data generated by EO satellites and other sources:

- CEOS-chair: NASA
- Homepage: http://ceos.org
- Members: 61 Agencies operating 192 satellites. Examples are: USGS, NASA, and NOAA (US), ESA (Europe), Chinese Academy of Space Technology (CAST), Japan Aerospace Exploration Agency (JAXA) (Japan), United Kingdom Space Agency (UKSA) (UK), Centre National d'Études Spaciales (CNES) (France), Canadian Space Agency (CSA) (Canada), Roskosmos (Russia), Deutsches Zentrum für Luft- und Raumfahrt (DLR) (Germany), Food and Agriculture Organization (FAO, UN), and ISPRS (photogrammetry and remote sensing).

Defence Geospatial Information Working Group (DGIWG)

The DGIWG represents the experts on digital geographic information in the military agencies primarily of the NATO countries. Established in 1983, the DGIWG has developed the digital geographic information exchange standard (DIGEST). However today, DGIWG primarily adopts or profiles ISO base standards.

ISO/TC 211 and DGIWG have signed a cooperative agreement according to which DGIWG publications or NATO standardization agreements (STANAGs) will be aligned with ISO or IEC deliverables:

- Members: Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Estonia, France, Germany, Greece, Italy, The Netherlands, New Zealand, Norway, Portugal, Romania, South Africa, Spain, Sweden, Turkey, UK, USA; Associates: Finland, Latvia, Poland
- Homepage: https://www.dgiwg.org.

Energistics

Energistics provides a noncompetitive, vendor-neutral infrastructure for energy industry professionals to do the work required to develop, maintain, adopt, and deploy standards. The predecessor of Energistics was formed in October 1990 by five founding sponsor oil companies: BP, Chevron, Elf (since merged into TotalEnergies), Mobil (since merged into ExxonMobil), and Texaco (since merged into Chevron) under the name Petrotechnical Open Software Corporation (POSC):

- Members: over 100 global exploration and production technology and oilfield service companies
- Chair: UK
- Homepage: http://www.energistics.org.

European Space Agency (ESA)

ESA's task is to draw up the European space programme and carry it through: to find out more about Earth, its immediate space environment, our Solar System and the Universe, as well as develop satellite-based technologies and services:

- Objectives: back in 1975, the ESA was founded by combining the activities of earlier and smaller organizations that had been building the first European space activities since 1960.
- Members: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Luxembourg, The Netherlands, Norway, Poland, Portugal, Romania, Spain, Sweden, Switzerland, and the United Kingdom.
- Associate member: Slovenia, Latvia, Lithuania.
- Cooperation agreement: Bulgaria, Canada, Croatia, Cyprus, Malta, Slovakia.
- Chair: Austria; headquarters: Paris.
- Homepage: http://www.esa.int.

European Umbrella Organization for Geographic Information (EUROGI)

Back in 1993, EUROGI was promoted by the Directorate General XIII (Information Society) of the European Commission. EUROGI facilitates value creation from location-based Information for a sustainable, prosperous, and cohesive Europe, and represents users and providers from the public, academic, and private sectors:

- Members: about 20 European Companies in the field of geographic information.
- Chair: Belgium.
- Homepage: http://eurogi.org/.

European Spatial Data Research (EuroSDR)

EuroSDR is a European organization linking national mapping and cadastral agencies (NMCAs) with research institutes and universities for the purpose of applied research with a focus on photogrammetry and other image-based technologies. EuroSDR was founded in 1953 in Paris as the Organisation Européenne d'Études Photogrammétriques Expérimentales (OEEPE) in the frame of an international treaty:

- Scope: EuroSDR is active in many research fields. Some of them are: new platforms for remote sensing, camera certification, image radiometry, mobile mapping, imagery and lidar integration, virtual globes, data archiving, generalization, standards, INSPIRE implementation, image matching, and land cover.
- Meetings: Twice a year.
- Members: 20 member states, each with a representative from an NMCA and a research organization.
- Chair: Germany.
- Homepage: http://www.eurosdr.net/.

The Food and Agriculture Organization of the United Nations (FAO/UN)

The Food and Agriculture Organization of the United Nations (FAO) was founded in 1945 with a mandate to raise levels of nutrition and standards of living, to improve agricultural productivity, and to improve the condition of rural populations. Today, the FAO is one of the largest specialized agencies in the United Nations system and the lead agency for agriculture, forestry, fisheries, and rural development.

FAO's mandate is to support members in their efforts to ensure that people have regular access to enough high-quality food. FAO can help by supporting policies and political commitments that promote food security and good nutrition and by making sure that up-to-date information about hunger and malnutrition challenges and solutions is available and accessible:

- Members: 195 member countries.
- Chair: USA; headquarters in Rome, Italy.
- Homepage: http://www.fao.org.

International Federation of Surveyors (FIG)

Founded in 1878 in Paris, FIG is the premier international organization representing the interests of surveyors worldwide.

Surveyors are professional people enabled to advice on the management and use of both rural and urban land and property, whether developed or undeveloped. Surveyors understand the legislation governing land and property, markets, and the economics of construction, management, maintenance, acquisition, and disposal:

- Members: more than 120 countries represented (associations, affiliates, corporate, academics).
- Meetings: annual working weeks and a congress every 4 years.
- Chair: Germany.
- Homepage: http://www.fig.net.

IEEE Geoscience and Remote Sensing Society (GRSS)

The fields of interest of the society are the theory, concepts, and techniques of science and engineering as they apply to the remote sensing of the Earth, oceans, atmosphere, and space, as well as the processing, interpretation, and dissemination of this information.

The members of GRSS come from both engineering and scientific disciplines. Those with engineering backgrounds are often familiar with geophysics, geology, hydrology, oceanography, and/or meteorology. The fusion of geoscientific and engineering disciplines in projects of global scope give the GRSS a unique interdisciplinary character. The society was first known as the Geoscience Electronics Group, formed in 1962:

- Chair: USA.
- Meetings: Since 1981 it has sponsored the annual International Geoscience and Remote Sensing Symposium (IGARSS) series.
- Homepage: http://www.grss-ieee.org/.

International Association of Geodesy (IAG)

The International Union of Geodesy and Geophysics (IUGG) comprises eight semiautonomous associations, one of them being the International Association of Geodesy (IAG).

Geodesy measures the figure of the Earth. This comprises the establishment of reference systems, monitoring the gravity field and rotation of the Earth, the deformation of the Earth's surface including ocean and ice, and positioning for interdisciplinary use:

- Chair: Finland.
- Homepage: http://www.iugg.org/associations/iag.php.

International Cartographic Association (ICA)

ICA's mission is to promote the disciplines and professions of cartography and GIScience in an international context. The objectives are illustrated by the topics of the commissions: art and cartography, atlases, cartographic heritage, cartography and children, early warning and crisis management, cognitive issues, education and training, generalization and multiple representation, geospatial analysis and modeling, GI for sustainability, history of cartography, location-based services, map design, map production, map projections, maps and graphics for the blind and partially sighted people, maps and the Internet, mountain cartography, open-source geospatial technologies, planetary cartography, SDI and standards, sensor-driven mapping, topograhic mapping, toponymy, ubiquitous mapping, use, user, and usability issues, visual analytics:

- Members: more than 100 national and affiliate members (institutions, companies).
- Chair: USA.
- Homepage: http://www.icaci.org.

International Civil Aviation Organization (ICAO)

The ICAO is a UN specialized agency that deals with all international civil aviation questions. The membership includes practically all countries of the world. Aircraft navigation is the ICAO topic that generates the most interest in the standardization of geographic information.

ICAO works to reach consensus on international civil aviation standards and recommended practices (SARPs) and policies in support of a safe, efficient, secure, economically sustainable, and environmentally responsible civil aviation sector. The organization was founded in 1944 in Chicago. Its predecessor, the International Commission on Air Navigation (ICAN), was founded in 1910 in Paris:

- Members: 193 states.
- Meetings: Assembly meets at least once every 3 years.
- Chair: Italy; headquarters in Montreal, Canada.
- Homepage: http://www.icao.int.

International Hydrographic Organization (IHO)

The IHO is an organization that consists of hydrographic agencies from most maritime countries around the world. The IHO is an intergovernmental consultative and technical organization that was established in 1921 to support safety in navigation and protection of the marine environment. One of their major efforts is the creation of international standards for digital hydrographical charts. For hydrographic charts, in order to promote and coordinate the development of standards, specifications and guidelines for official products and services, the International Hydrographic Organization established a Hydrographic Services and Standards Committee

(HSSC) to monitor the requirements of mariners and other users of hydrographic products and information systems:

- Objectives:
  - Coordination of the activities of the national hydrographic offices
  - Greatest possible uniformity in nautical charts and documents
  - Adoption of reliable and efficient methods of carrying out hydrographic surveys
  - Development of the sciences in the field of hydrography and oceanography
- Members: about 90 maritime states
- Chair: Germany; International Hydrographic Bureau (IHB) in Monaco
- Homepage: http://www.iho.int/.

International Association of Oil and Gas Producers (IOGP)

The International Association of Oil and Gas Producers (IOGP) is the UK-based voice of the global upstream oil and gas industry. It functions to improve understanding of the oil and gas industry by being a visible, accessible, reliable, and credible source of information and to represent and advocate the views of this industry:

- Members: About 80 companies and national associations
- Chair: UK
- Homepage: http://www.iogp.org.

International Society for Photogrammetry and Remote Sensing (ISPRS)

Photogrammetry and remote sensing is the art, science, and technology of obtaining reliable information from noncontact imaging and other sensor systems about the Earth and its environment. Established in 1910, the ISPRS is a non-governmental organization devoted to the development of international cooperation for the advancement of photogrammetry and remote sensing and their applications.

The ISPRS was the traditional international society for photogrammetry, later extending its scope to include remote sensing. The society primarily represents the engineering point of view, gathering all research on sensors (including their models) and on the mathematical models for derivation of spatial information from imagery. The ISPRS offers expertise on three-dimensional geographic information:

- Members: 92 national societies (ordinary)
- Meetings: ISPRS congresses every 4 years, midterm symposia in between
- Chair: Germany
- Homepage: http://www.isprs.org.

Joint Research Centre of the European Union (JRC)

The JRC is the research center of the European Commission. The European Commission functions as the government of the European Union (EU). One of the six laboratories of the JRC governs the creation of the European spatial data infrastructure (INSPIRE) and is located in Ispra, Italy:

- Members: Directorate-General and 6 JRC Institutes
- Chair: Bulgaria; Directorate-General in Brussels, Belgium
- Homepage: http://ec.europa.eu/dgs/jrc/.

Object Management Group (OMG)

Founded in 1989, the OMG is a computer industry consortium that develops standards related to many IT fields.

OMG develops enterprise integration standards for a wide range of technologies, including: realtime, embedded and specialized systems, analysis and design, architecture-driven modernization, and middleware, business modeling and integration, finance, government, healthcare, legal compliance, life sciences research, manufacturing technology, robotics, software-based communications, and space:

- Members: about 330 companies and universities, representing virtually every large organization in this field, smaller companies, and end-users
- Chair: USA
- Homepage: http://www.omg.org.

Open Source Geospatial Foundation (OSGeo)

The Open Source Geospatial Foundation is a not-for-profit organization whose mission is to foster global adoption of open geospatial technology by being an inclusive software foundation devoted to an open philosophy and participatory community driven development.

- Goals
  - Provide resources for foundation projects
  - Foster the use of open source geospatial software
  - Encourage interoperability with open and community standards
  - Ensure interoperability between the foundation projects
  - Encourage a high degree of quality and innovation in foundation projects
  - Champion the use of open-source and community participation through the development of an open education curriculum
  - Enable communication and cooperation amongst OSGeo communities
  - Champion community building through horizontal (local chapters) engagement and through vertical (sector specific) agreements with like-minded organizations

  - Be a welcoming and inclusive worldwide organization at all levels
  - Celebrate excellence, openness and service within the OSGeo community.
- Members: about 500
- Chair: Greece
- Homepage: https://www.osgeo.org

Organization for the Advancement of Structured Information Standards (OASIS)

Founded in 1993, OASIS is a consortium that drives the development, convergence, and adoption of open standards for the global information society.

OASIS promotes industry consensus and produces worldwide standards for security, Internet of Things, cloud computing, energy, content technologies, emergency management, and other areas:

- Members: over 600 organizations and individual members in more than 65 countries
- Chair: USA
- Homepage: http://www.oasis-open.org.

Open Geospatial Consortium, Incorporated (OGC)

The Open Geospatial Consortium (OGC) is an international organization committed to making quality open standards for the global geospatial community.

- Members: about 550 (government, commercial organizations, NGOs, academic, and research institutes)
- Chair: USA
- Homepage: http://www.opengeospatial.org.

A detailed discussion of the OGC follows in Sect. 15.5.

Pan-American Institute of Geography and History (PAIGH)

Created in 1928 in Havana, Cuba, the Pan-American Institute of Geography and History (PAIGH) is a technical and scientific body of the Organization of American States specializing in the areas of cartography, geography, history, and geophysics. The organization has promoted collaboration throughout the Americas by sponsoring conferences, publishing journals, and organizing workshops:

- Members: most countries in Latin America plus some overseas, such as Spain
- Homepage: https://uspaigh.org.

Small Business Standards (SBS)

Established in 2013, Small Business Standards (SBS) is a European association cofinanced by the European Commission and EFTA Member States. SBS's goal is to represent and defend small and medium-sized enterprises' (SMEs) interests in the standardization process at European and international levels:

- Members: 20 associations of SMEs
- Chair: Sweden
- Homepage: http://www.sbs-sme.eu/.

Scientific Committee on Antarctic Research (SCAR)

Starting during the International Geophysical Year 1957–1958, SCAR coordinates scientific activities regarding the Antarctic.

SCAR has three Science Groups titled GeoScience, Life Science, and Physical Science. The Science Groups are subdivided into over 30 groups that address, for instance the Geodetic Infrastructure of Antarctica (GIANT) and the International Bathymetric Chart of the Southern Ocean (IBCSO):

- Members: about 30 national members and about 20 others (associated members and international unions)
- Chair: USA
- Homepage: http://www.scar.org.

United Nations Economic Commission for Africa (UN ECA)

In 1958, the United Nations (UN) established the Economic Commission for Africa (ECA) as one of the UN's five regional commissions. ECA's mandate is to promote the economic and social development of its member states, foster intraregional integration, and promote international cooperation for Africa's development:

- Objectives:
  - Macroeconomic policy
  - Regional integration and trade
  - Social development
  - Natural resources
  - Innovation and technology
  - Gender
  - Governance
  - Statistic
- Members: about 50 member states (practically all of Africa)
- Chair: Cameroon; headquarters in Addis Ababa, Ethiopia
- Homepage: http://www.uneca.org/.

United Nations Economical Commission for Europe (UNECE)

UNECE's major aim is to promote pan-European economic integration, and it is one of the UN's five regional commissions.

As a multilateral platform, UNECE facilitates greater economic integration and cooperation among its member countries and promotes sustainable development and economic prosperity:

- Members: 56 member states (Europe plus North America and Central Asia)
- Chair: Slovakia; headquarters in Geneva, Switzerland
- Homepage: http://www.unece.org/.

United Nations Group of Experts on Geographical Names (UNGEGN)

The UNGEGN is a UN expert group that provides technical recommendations for the international spelling of geographic names. In 1959, the Economic and Social Council (ECOSOC) paved the way for a small group of experts to meet and provide technical recommendations on standardizing geographical names at the national and international levels:

- Chair: USA
- Homepage: http://unstats.un.org/unsd/geoinfo/ungegn/.

United Nations Global Geospatial Information Management (UN-GGIM)

In 2011, the Economic and Social Council (ECOSOC) of the United Nations created the UN Committee of Experts on Global Geospatial Information Management (UN-GGIM). It is a subdivision of the UN Statistics Division.

The areas of work include the development of the global geodetic reference frame, the adoption and implementation of standards, legal and policy frameworks including critical issues related to authoritative data, and determining fundamental data sets:

- Members: almost 200 partner states
- Chair: Germany
- Homepage: http://ggim.un.org/.

Regional Committee of the UN Global Geospatial Information Management Americas (UN-GGIM: Americas)

- Objectives:
  - To establish and coordinate policies and technical standards for the development of regional geospatial data infrastructure in the Americas.
  - To promote as a matter of priority the establishment and development of the national geospatial data infrastructures of each of the members of the UN-GGIM Americas.
  - To prompt the exchange of geospatial information among all members of the Americas community, respecting their autonomy, in accordance with their national laws and policies.
  - To encourage cooperation, research, complementation and exchange experiences in areas of knowledge related to the geospatial subject.
  - To define guidelines and strategies to support member nations in the development of cadastral information taking into account the individual needs of each country.
- Members: 36 countries in North and South America
- Chair: Mexico
- Homepage: http://www.un-ggim-americas.org/en/.

United Nations Global Geospatial Information Management for Asia and the Pacific (UN-GGIM-AP)

- The United Nations Global Geospatial Information Management for Asia and the Pacific (UN-GGIM-AP) is a regional committee. It focuses its activities on geodetic reference framework, disaster risk management, regional SDI, as well as cadastre and land management.
- Members: 56 countries in Asia and the Pacific
- Chair: Australia
- Homepage: http://www.un-ggim-ap.org.

Universal Postal Union (UPU)

Proposed in Germany and established in 1874, the Universal Postal Union (UPU, German: Weltpostverein), with its headquarters in Bern, Switzerland, is the second oldest international organization worldwide.

The UPU helps to ensure a universal network of up-to-date products and services. This way, the organization fulfills an advisory, mediating, and liaison role, and provides technical assistance where needed. It sets the rules for international mail exchanges and makes recommendations to stimulate growth in mail, parcel, and financial service volumes and improve quality of service for customers:

- Member: 193 member countries
- Chair: Ivory Coast; headquarters in Bern, Switzerland
- Homepage: http://www.upu.int/.

World Wide Web Consortium (W3C)

The World Wide Web Consortium is an international community that develops open standards to ensure the long-term growth of the Web.

- Principles: Web for all, Web on everything
- Vision: Web of consumers and authors, data and services, trust
- Chair: UK
- Homepage: https://www.w3.org.

World Geospatial Industry Council (WGIC)

- Overview
  The World Geospatial Industry Council is an association of companies representing the entire ecosystem of geospatial industry. WGIC endeavors to:
  1. Enhance the role of the geospatial industry and strengthen its contribution in global economy and society.
  2. Facilitate exchange of knowledge within the geospatial industry and co-creation of larger business opportunities for the geospatial industry.
  3. Represent business interest, share perspectives of the geospatial industry and undertake policy advocacy and dialogue with public authorities, multilateral agencies and other relevant bodies.
- Members: A Membership into WGIC is open only for commercial entities.
- Chair: USA (Esri)
- Homepage: https://wgicouncil.org.

The World Meteorological Organization (WMO)

The WMO is the worldwide weather organization dealing with weather prediction, climate change, and related topics.

From weather prediction to air pollution research, climate change-related activities, tropical storm forecasting, and water-related hazards, the WMO coordinates global scientific activity to allow increasingly prompt and accurate weather information and other services for public, private, and commercial use, including international airline and shipping industries:

- Members: 191 states and territories
- Chair: Finland; headquarters in Geneva, Switzerland
- Homepage: http://www.wmo.int/.

World Bank

Founded in 1944 during the Bretten Woods conference and led by the USA since, the World Bank is committed to reducing extreme poverty and works as a large source of funding and knowledge:

- Objectives: the World Bank is subdivided in five institutions.
  – The International Bank for Reconstruction and Development
  – The International Development Association
  – The International Finance Corporation
  – The Multilateral Investment Guarantee Agency
  – The International Center for Settlement of Investment Disputes
- Members: 189 countries
- Chair: USA
- Homepage: http://www.worldbank.org/.

15.5 Open Geospatial Consortium

The Open Geospatial Consortium (OGC) is an international industry consortium dedicated to geospatial interoperability by instituting international consensus-based standards.

Its mission is

> To advance the development and use of international standards and supporting services that promote geospatial interoperability. To accomplish this mission, OGC serves as the global forum for the collaboration of geospatial data/solution providers and users.

15.5.1 Background

The spirit and development process are different from the ISO approach in that the OGC primarily develops implementation specifications, whereas the ISO standardizes abstract specifications at the top level.

Founded in 1994, the OGC evolved from the Geographic Resources Analysis Support System (GRASS) user community – the Open GRASS Foundation (OGF). GRASS was developed by the US Army Corps of Engineers Construction Engineering Research Laboratory (CERL) and used by several US Government agencies and universities around the world. As more and diverse GIS systems evolved, it was realized that the open/data format espoused by the OGF was not providing an inclusive interoperable solution. On September 25, 1994 the OGC was founded to provide full interoperability for diverse geoprocessing systems communicating directly over networks using open interfaces based on the open geodata interoperability specification.

Today, the OGC has approximately 550 members, representing the complete spectrum of players in the GIS marketplace from around the world (Fig. 15.47). The member-

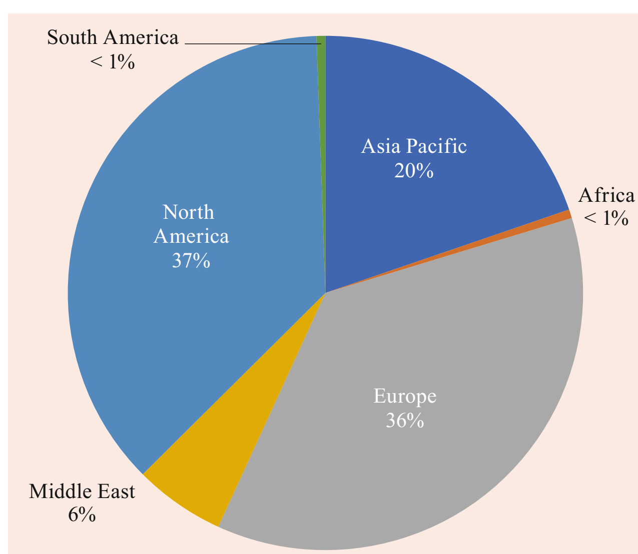


Fig. 15.47 OGC global membership distribution (as of Nov 2021)

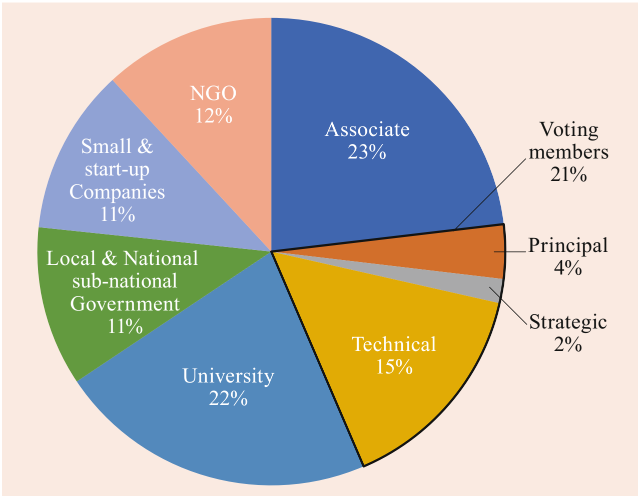


Fig. 15.48 OGC membership by level (as of Nov 2021)

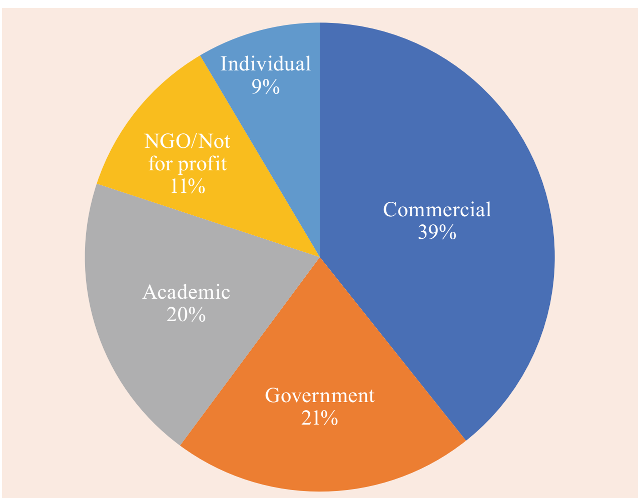


Fig. 15.49 OGC membership by sector (as of Nov 2021)

ship type is split into four main categories: strategic, principle, technical, and associate, (with four discipline related categories Fig. 15.48). Their background ranges from GIS manufacturers to geographic information production companies, government agencies, universities and research laboratories (Fig. 15.49). The annual fee is different for each category: strategic members US $250 000 (including fees and in-kind contributions), principal members US $60 000, technical members US $12 000, and associate members between US $550 and US $4800, e.g., US $550 in the case of universities, to US $250 for municipal government members. In practice, the annual fee of strategic members is negotiable, because the fee is often paid by installing a working position at the staff level and/or by sponsoring larger implementation projects.

The OGC is registered as a not-for-profit corporation under Section 501(a) of the US Internal Revenue Code. It has approximately 26 officers and staff. Nadine Alameh is

Table 15.9 List of strategic members of the open geospatial consortium (as of November 2021)

Company	Field of activity	Origin
European Space Agency (ESA)	Government/ civilian	Europe
GeoConnections – Natural Resources Canada	Government/ civilian	North America
UK Ordnance Survey	Government/ civilian	Europe
US Department of Homeland Security (DHS)	Government/ civilian	North America
United States Geological Survey (USGS)	Government/ civilian	North America
US National Aeronautics and Space Administration (NASA)	Government/ civilian	North America
US National Geospatial-Intelligence Agency (NGA)	Government/ defense	North America

the Chief Executive Officer (CEO). The Board of Directors (BOD) consists of 17 people from around the world, representing a variety of disciplines.

The OGC is managed by a hierarchical set of committees and boards governed by the CEO and the BOD (Fig. 15.50):

1. The Strategic Member Advisory Committee consists of the strategic members (Table 15.9) and the OGC CEO. It recommends areas of strategic opportunity and resource strategies to support OGC program operations to the BOD, the OGC staff, and the membership. It also provides the innovation and specification programs with management and operational resources.
2. The planning committee is made up of strategic and principal members (Table 15.10), the executive committee of the BOD, and two representatives of the technical committee, and is chaired by an OGC staff member. The planning committee performs strategic technology planning with regard to the geospatial standards that have the greatest chance of being adopted by the market. It ratifies specification development plans, release schedules, conformance and testing plans, and all major documents produced by the technical committee. It also elects the BOD and maintains the policy and procedures of the consortium.
3. The technical committee (TC) is made up of strategic, principal, and technical members, and is chaired by OGC staff. This is the committee where the standards are developed through a consensus process. All levels of membership can participate and provide input to developing documents, but Technical Committee members are the only members allowed to vote to approve standards and other major documents to be forwarded to the planning committee for final approval. The technical committee establishes the special interest groups, working groups, and subcommittees to perform the work of the consortium.
4. The OGC Architecture Board (OAB) is facilitated by the OGC Chief Technical Officer and one additional OGC

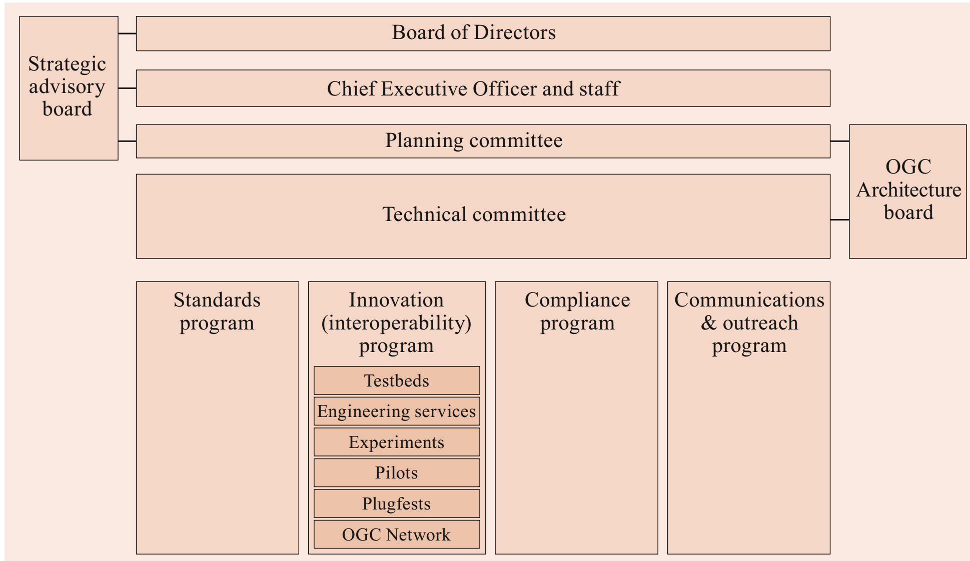


Fig. 15.50 OGC organization

Table 15.10 List of principal members of the open geospatial consortium (as of November 2021)

Company	Field of activity	Origin
Airbus Defence and Space	GIS products/services	Europe
Amazon Web Services, Inc.	Information technology	USA
Cubic Mission and Performance Solutions – Digital Intelligence	GIS products/services	USA
Defence Science & Technology Laboratories (Dstl)	Government/defense	Europe
Department of Science and Technology	Government – civilian	India
Esri	GIS products/services	USA
Feng Chia University	University	Taiwan
Google	Information technology	USA
Hexagon	GIS products/services	USA
Leidos	GIS services	USA
Maxar	GIS products/services	USA
Microsoft Corporation	GIS products/services	USA
Oracle USA	Information technology	USA
Trimble Navigation Ltd.	GIS products/services	USA
United Kingdom Hydrographic Office	Government/civilian	Europe
US Army Geospatial Center	Government/defense	USA
US Census Bureau	Government/civilian	USA
US National Oceanic and Atmospheric Administration (NOAA)	Government/civilian	USA

staff and is made up of 15 OGC technical committee representatives, nominated and elected by the technical committee. The OAB provides guidance to the TC and PC regarding OGC standards lifecycle management. It reviews and recommends for adoption by the TC the OGC reference model, monitors technology trends related to the standards baseline, and identifies technology gaps and issues. It reviews all RFC submissions of candidate standards and provides guidance related to the technical content of existing OGC standards and best practices [97].

15.5.2 OGC Programs

The work in OGC is performed in four programs with different objectives (Fig. 15.50).

The Standards Program

The Standards Program is where the planning and technical committee and its working groups develop and approve the OGC standards and major documents. Its processes are described under the OGC standards development process below (Sect. 15.5.3).

The OGC Innovation Program

The Innovation Program (IP) [98] is a series of engineering initiatives, including *OGC Testbeds*, *Engineering Services, Interoperability Experiments*, *Interoperability Pilots* and *Plugfests*, aimed at accelerating the development and acceptance of OGC specifications. The innovation program evolved into the OGC's main field of activity.

The *OGC Testbed* is a research and development activity to determine specification requirements and proofs of concept. Typically, sponsors with requirements work with the OGC IP team to develop a concept to determine and test specification requirements. This team develops and puts out a request for quotation/call for participation. Potential participants submit proposals, which are then reviewed, and participants are selected. Participants may be provided with partial funding but are expected to also provide some in-kind contribution of resources. Some participants participate fully in-kind. The testbed usually follows an architecture developed by the sponsors and the OGC IP team. Prototype proof-of-concept applications are developed. The outcome of these testbeds is a demonstration of concepts, what worked and what did not, and which specifications are required, as well as engineering reports (ERs), which are presented in the domain working groups (DWGs) and some of which are submitted through the request for comment (RFC)/standards working group (SWG) process to become standards.

Engineering Services are where OGC staff provide contracted services assisting organizations with how to apply OGC standards and an interoperability architecture to fit an organization's business needs. Types of services offered are:

- *Assessment, analysis, and feasibility support,* which can include requirement's definition, business plans, use-case modeling, analysis, and procurement assessments.
- *Standards-based reference architecture design and documentation* are services that provide recommendations to improve an organization's interoperability using well-known reference model-based viewpoints to provide standards-based architectures.
- *Operation and/or enhancement of OGC online compliance facilities and compliance evaluation* are services in which OGC supports members with their compliance testing, identifying compliance deficiencies in an organization or its product portfolio.
- *Education and outreach* are services that include interoperability training for organizations, presentations at conferences and symposiums, and participation in public relations and programs related to the uptake of open standards.
- *Innovation and research services*; OGC provides assessment, analysis, and feasibility support to research and development programs to enhance the OGC technology baseline and experiment with new technologies aimed at positioning OGC as an innovation hub within the international research and development community.

Interoperability experiments (IEs) are an activity to further improve existing OGC standards, led and executed by OGC members and usually focusing on a specific domain, facilitated by OGC staff. Any member of the OGC can participate; all participation is in-kind; there are no sponsored activities. These are usually short-duration activities, narrowly focusing on a single interoperability issue. An example of an IE is the geo-interface for atmosphere, land, Earth, and ocean netCDF (GALEON), which experimented with providing a geo-interface to netCDF datasets using the OGC WCS specification to provide open access to atmospheric and oceanographic modeling and simulation outputs. It also provided a test and comments/suggestions for future versions of the WCS specification. The initiators of the interoperability experiment were: the Unidata/University Corporation for Atmospheric Research (UCAR), the Institute of Methodologies for Environmental Analysis of the Italian National Research Council (IMAA-CNR), the George Mason University, and the NASA Geospatial Interoperability Office.

Pilots are run much like testbeds, except instead of developing and testing technology for new standards, they stress-test and perfect existing OGC standards or sets of standards, working toward improving the standard. Like testbeds, pilots have sponsors and issue requests for quotation/calls for participation (RFQ/CFP) to select applicants as participants who accept partial funding for their efforts. An example OGC pilot is the group on earth observation system of system (GEOSS) architecture implementation pilot (AIP), which was an interoperability program to test OGC catalogue services specifications and improve access to data and services, as well as to establish and test infrastructure components and OGC services for several GEOSS societal benefit areas, such as disaster management, biodiversity and climate change, air quality and health, and renewable energy. GEO members and participating organizations provided components and services relevant to SBAs. They also participated in interoperability testing of the services to validate the architecture and in the collaborative refinement of societal benefit scenarios to guide testing, demonstrations, and operations of interoperable services. The pilot was initiated by the group on Earth observations (GEO), a voluntary partnership of 148 governments and international organizations launched in response to calls for action by the 2002 World Summit on Sustainable Development and by the G8 (Group of Eight) leading industrialized countries.

Plugfests provide a setting in which GIS software developers and others can interoperate – testing the interoperability of OGC standards.

The Compliance Program

This program provides a testing facility for software implementors to test and certify their products compliant to OGC standards. This allows software developers to be assured their products are compliant to OGC standards and for customers to be able to shop for compliant products. It provides a free online testing facility and a process for the certification of compliant products. It maintains an online list of certified products by OGC standard/version of standards (http://www.opengeospatial.org/resource/products/compliant). Although the online tests are free, permission to use the "Certified OGC Compliant" Mark associated with each version of an OGC standard organizations must pay an annual fee determined by their membership and level of membership in OGC according to their organization's total gross annual revenue.

15.5.3 The OGC Standards Development Process

The OGC has established a formal process for the development of standards and major documents. Work is performed in the technical committee and approved by the planning committee using a defined consensus-based voting procedure and a number of formalized document types, including implementation specification (IS), best practice paper (BP), engineering report (ER), discussion paper (DP), white paper (WP), and request for comment (RFC). Only IS are officially approved OGC standards to be used to specify requirements in contracts and other official documents. A BP is an official position of the OGC endorsed by the membership and should not be used for contracting purposes. An ER is the output of an OGC innovation program testbed and does not represent the official position of the OGC. They may contain specifications and requirements that often provide the foundation for the development of an OGC IS. An ER may also contain a report on the testbed results and a summary of the outcome. DPs are used to provide opinions to create discussion around technology issues and as such do not represent the official position of the OGC. WPs, however, do provide official positions of the OGC, but should not be treated as standards.

The work of the TC is performed in working groups. The Standards Working Groups (SWGs) develop draft specifications, which are then approved by the technical committee membership. The Domain Working Groups (DWGs) focus on domain-specific requirements and issue and review ERs and encourage or produce RFCs for new candidate standards. The DWG submits a document, intended to become a standard, to the TC, which approves the formation of a SWG, where work on the standard is performed. Access to work in the SWG is limited to individuals who have to *opt in* to repudiate intellectual property right claims to the future standard [99]. Voting in WGs is by simple majority of OGC members present at the WG meeting, with the caveat that no OGC member organization may cast more than one vote. Once the work of the SWG is finished, the candidate standard is released to the TC for a 60-day intellectual property rights (IPR) review and vote. Once approved by the TC, it is sent to the PC for final vote and publishing as an IS. OGC also has standards known as "Community Standards". These are standards developed outside of OGC with strong evidence of implementation. Community standards are considered normative OGC standards, an official position of the OGC. The original developers share the IPR with OGC or grant unlimited free use of intellectual property to all implementers.

Presently, the OGC has the DWGs and SWGs (Table 15.11).

15.5.4 OGC Standards

The vision of the OGC is a world in which everyone benefits from geographic information and services that are made available across any network, application, or platform. Through the tremendous efforts of the OGC in the past, some of the visionary developments, such as Web mapping, have made considerable progress. OGC standards fall into two categories: abstract specifications and implementation specifications. Originally, the concept of the OGC specification program was aimed at building a complete suite of GIS standards. Today, this ambitious goal has been unofficially modified towards a focus on implementation specifications. The work on the abstract specification has become the concern of ISO/TC 211 *Geographic information/Geomatics*. Some of the original OGC abstract specifications have become ISO standards, and some ISO standards have become OGC abstract specifications.

OGC Abstract Standards

OGC abstract standards provide the conceptual foundation for OGC standards. By basing OGC implementation standards on an architecture of abstract standards that have been harmonized with other standards in the fields of information technology and geographic information such as the W3C and ISO/TC 211 ensures proper reuse, integration, and harmonization within the OGC implementation standards and those of the other standards bodies (Table 15.12).

Topic 1: Features and Geometry – Part 1: Feature Models

This abstract standard [100] describes how feature models are used to create digital entities to represent real world phenomena. It defines how a feature model is structured, created, stored, queried, and manipulated. The previous version of this standard used static object models to provide conceptual schemas for describing the spatial characteristics of geographic features, as well as a set of spatial operations consistent with those schemas. This standard covers dynamic schematic, onotological, and taxonomic definition representations of features and their properties and relations. It provides definitions of the structures and operations associated to these digital entities to represent, manipulate, and query feature data based on those models.

Topic 2: Spatial Referencing by Coordinates

This abstract specification [101] defines the conceptual schema for the description of spatial referencing by coordinates, optionally extended to spatiotemporal referencing. It describes the minimum data required to define one, two, and three-dimensional spatial coordinate reference systems with an extension to merged spatiotemporal reference systems.

Table 15.11 Domain working groups (DWGs) and working groups (WGs) of the OGC (as of November 2021)

Title	Chair
3DIM DWG	Ordnance Survey CAE Inc. National University of Singapore
3D GeoVolumes SWG	
3-D Portrayal SWG	LiRiS Hochschule für Technik, Stuttgart
Agriculture DWG	George Mason University Plan4All z.s.
Architecture DWG	Universitat Autònoma de Barcelona (CREAF)
Artificial Intelligence in Geoinformatics DWG	Feng Chia University Dimitris Kotzinos National Institute of Advanced Industrial Science & Technology (AIST) Anno.ai Helyx secure information systems Ltd.
Aviation DWG	NASA EUROCONTROL
Big Data DWG	Oracle Jacobs University Bremen GmbH Charles Heazel
Blockchain and Distributed Ledger Technologies DWG	Secure Dimensions GmbH ESA
CDB SWG	University of Calgary CAE Inc.
Citizen Science DWG	Universitat Autònoma de Barcelona (CREAF) PSMA Australia Woodrow Wilson International Center for Scholars
Coordinate reference system DWG	Esri UK Met Office
CityGML SWG	Open Site Plan virtualcitySYSTEMS GmbH
Coverages DWG	Jacobs University Bremen GmbH EOX IT Services GmbH Ecere Corporation
Coverages SWG	Jacobs University Bremen GmbH EOX IT Services GmbH Ecere Corporation
CRS SWG	Esri International Association of Oil and Gas Producers (IOGP)
CRS well known text SWG	Esri International Association of Oil and Gas Producers (IOGP)
Data preservation DWG	Universitat Autònoma de Barcelona (CREAF)
Data quality DWG	Universitat Autònoma de Barcelona (CREAF) Matt Beare Helyx secure information systems Ltd Curtin University
Discrete global grid systems DWG	Matthew Purss Joint Research Centre (JRC)
Discrete global grid systems SWG	Landcare Research New Zealand Limited Matthew Purss
Defense and intelligence DWG	European Union Satellite Centre US National Geospatial-Intelligence Agency (NGA)
Earth Observation Exploitation Platform DWG	European Space Agency (ESA) Terradue Srl GeoConnections – Natural Resources Canada
Earth systems science DWG	Joint Research Centre (JRC)
Emergency & disaster management DWG	US National Aeronautics and Space Administration (NASA) US Dept of Defense/DISA Feng Chia University UK Met Office

Table 15.11 (Continued)

Title	Chair
Energy and utilities DWG	GeoConnections – Natural Resources Canada Eddie Oldfield
Environmental Data Retrieval API SWG	UK Met Office Wuhan University US Geological Survey (USGS)
EO product metadata and OpenSearch SWG	con terra GmbH Spacebel s. a.
Features and Geometries JSON SWG	CubeWerx interactive instruments GmbH US Army Geospatial Center Geonovum
Features API SWG	CubeWerx interactive instruments GmbH
GeoAPI SWG	GEOMATYS
Geocoding API SWG	PSMA Australia
GeoPackage SWG	Tracey Birch Image Matters LLC
GeoPose SWG	Ordnance Survey Open AR Cloud Association
GeoSciMLSWG	GeoConnections – Natural Resources Canada
GeoSPARQL SWG	CubeWerx PSMA Australia Dimitris Kotzinos
Geospatial user feedback SWG	Universitat Autònoma de Barcelona (CREAF)
Geoscience DWG	BRGM
GML DWG	interactive instruments GmbH
GML 3.3 SWG	interactive instruments GmbH
Geosemantics DWG	CSIRO Geonovum PSMA Australia
GeoSynchronization SWG	CubeWerx US Army Geospatial Center
GeoTIFF SWG	Universitat Autònoma de Barcelona (CREAF) Charles Heazel
GeoXACML SWG	Secure Dimensions GmbH
GMLJP2 SWG	European Union Satellite Centre
Groundwater SWG	GeoConnections – Natural Resources Canada
Health DWG	Health Solutions Rescearch, Inc.
HDF SWG	The HDF Group
Hydrology DWG	National Computational Infrastructure US Geological Survey (USGS) World Meteorological Organization (WMO)
Hydrologic features SWG	US Geological Survey (USGS)
IndoorGML SWG	UK Ordnance Survey Pusan National University
Intreroperable simulation and gaming DWG	CAE Inc. Leidos Department of Defence (Australia)
KML 2.3 SWG	Google
Land administration DWG	Delft University of Technology University of Melbourne Kadaster International
Land and infrastructure DWG	Trimble Navigation Ltd.
Land and infrastructure SWG	Hexagon Trimble Navigation Ltd.
Marine DWG	IIC Technologies Limited Teledyne Geospatial Caris and Optech US National Geospatial-Intelligence Agency (NGA) United Kingdom Hydrographic Office

Table 15.11 (Continued)

Title	Chair
Metadata and catalog DWG	Hexagon Katholieke Universiteit Leuven Byron Cochrane Agentschap Informatie Vlaanderen (AIV)
Meteorology and oceanography DWG	UK Met Office US National Oceanic and Atmospheric Administration (NOAA)
Mobile location services DWG	US Army Geospatial Center Pusan National University
Moving features SWG	Hitachi, Ltd., Defense Systems Company National Institute of Advanced Industrial Science and Technology (AIST) Université Libre de Bruxelles (ULB)
MUDDI SWG	Ordnance Survey United Kingdom Research and Innovation (UKRI) New York City Geospatial Information System and Mapping Organization (GISMO)
NetCDF SWG	University Corporation for Atmospheric Research (UCAR) Joint Research Centre (JRC)
O&M SWG	Spatineo Oy BRGM Katharina Schleidt
OGC API – Common SWG	Universitat Autònoma de Barcelona (CREAF) Charles Heazel
OGC API – Maps SWG	Universitat Autònoma de Barcelona (CREAF) Esri Science Systems & Applications, Inc. (SSAI)
OGC API – Processes SWG	52° North Spatial Information Research GmbH Hexagon Airbus Defence & Space
OGC API – Records SWG	CubeWerx Image Matters LLC Environment and Climate Change Canada
OGC API – Styles SWG	US Army Geospatial Center CubeWerx US Army Geospatial Center Strategic Alliance Consulting Inc.
OGC API – Tiles SWG	Universitat Autònoma de Barcelona (CREAF) Ecere Corporation
OWS Common – Security SWG	Secure Dimensions GmbH
OWS Context SWG	Envitia Ltd. Terradue Srl
Perspective imagery DWG	Danish Agency for Data Supply & Efficiency (SDFE) Eric Hirschorn
PipelineML SWG	Enterprise Products Merkator NV/SA
Point cloud DWG	Delft University of Technology Hexagon University College London
Point of interest SWG	Open AR Cloud Association Matthew Purss AfriGIS (Pty) Ltd.
Portrayal DWG	Esri Strategic Alliance Consulting Inc. Defence Science & Technology Laboratories (Dstl)
PubSub SWG	CNR Institute for Atmospheric Pollution Research
Quality of service and experience DWG	Environment and Climate Change Canada Spatineo Oy
Routing SWG	US Army Geospatial Center
Security DWG	Secure Dimensions GmbH

Table 15.11 (Continued)

Title	Chair
Sensor model language 2.0 SWG	Botts Innovative Research Airbus Defence & Space
SensorThings SWG	keys University of Calgary Fraunhofer-Gesellschaft
Sensor Web Enablement DWG	Botts Innovative Research University of Calgary
Simple features SWG	Oracle
Smart cities DWG	Oracle Trimble Navigation, Ltd Institut National de l'Information Géographique et forestière (IGN) PSMA Australia
Spatial data on the Web Working Group	UK Met Office Geonovum
Statistical DWG	US Census Bureau
Styles and Symbology Encoding SWG	University of Applied Sciences, Western Switzerland School of Business & Engineering Vaud (HEIG-VD) Lab-STICC CNRS UMR 6285 Ecere Corporation
Temporal DWG	Jacobs University Bremen GmbH UK Met Office
Temporal WKT for calendars SWG	UK Met Office
TimeSeriesML SWG	US National Oceanic and Atmospheric Administration (NOAA) UK Met Office
Training Data Markup Language for AI SWG	
Uninhabited Systems (UxS) DWG	US National Aeronautics and Space Administration (NASA) keys
University DWG	University of Calgary Feng Chia University University of Tartu
WaterML 2.0 SWG	
Workflow DWG	Hexagon

Table 15.12 OGC abstract standard – ISO equivalent standards

OGC abstract standard	ISO international standard
Topic 1 Feature geometry	ISO 19107 *Spatial schema*
Topic 2 Spatial referencing by coordinates	ISO 19111 *Spatial referencing by coordinates*
Topic 6 Schema for coverage geometry and functions	ISO 19123 *Schema for coverage geometry and functions*
Topic 7 Earth imagery	ISO/TS 19101-2 *Reference Model – Part 2: Imagery*
Topic 11 Metadata	ISO 19115-1 *Metadata*
Topic 12 The OGC service architecture	ISO 19119 *Services*
Topic 20 Observations and measurements	ISO 19156 *Observations and measurements*
Topic 21 Discrete Global Grid Systems	ISO 19170-1 *Discrete Global Grid Systems Specification*

It allows additional descriptive information to be provided. It also describes the information required to change coordinates from one coordinate reference system to another. It was developed in collaboration with ISO 19111.

Topic 3: Locational Geometry Structures

Locational geometry [102] provides essential and abstract models for technology that is used widely across the GIS landscape. Its first heavy use is in support of simple feature geometry specifications and their spatial reference systems. It provides a discussion of the notion of locational geometry. The scenario of this discussion assumes that the same project world has been (or is to be) implemented (that is, abstracted into a feature collection) twice, using two different locational systems. A locational system is a mathematical construct providing coordinates for each corner of interest. The coordinates are usually scalars, but can be values from another domains. Examples are provided in Table 15.13.

Topic 4: Stored Functions and Interpolation

This topic [103] provides essential and abstract models in support of coverage specifications. Coverages, in general, require two stored functions. The first relates Earth's coordinates to the window coordinates in the coverage extent, providing a mapping from the coordinates of a spatial refer-

Table 15.13 Coordinate location system examples

Locational coordinates	Meaning of coordinates
(x, y, z)	Where x, y, and (optionally) z are real numbers (abstract geometry coordinates)
(lon, lat, elev)	Where here longitude, latitude, and (optionally) elevation are geographic coordinates (world coordinates)
(n, x)	Where n is a segment ID and x is the linear offset along the segment from the origin of the segment (linear reference coordinates)
(r, c)	Where r and c are (perhaps integer or real) row and column coordinates (image or raster coordinates)
(E, N)	Where E and N (Easting and Northing) are real numbers (map coordinates)

ence system (SRS) to the coverage extent coordinates. The second function assigns values to points in the coverage extent. The values may be thought of as colors, as this is a common value space. However, the values could be temperatures, a scalar representing fitness of habitat for songbirds, the name of the owner of the parcel containing the point, and so on, defined by a schema and taking the schema mapping and taking values in the schema range.

Topic 5: Features

This abstract standard [104] introduces the concept and discusses the essential model for features that have been defined in ISO 19101 and an extensive primer on the notion of geographic information. It defines the abstract model for feature, feature identifier, identifier scope, identifier change registry, feature repository, and feature collection.

Topic 6: Schema for Coverage Geometry and Functions

This abstract standard [105] defines a conceptual schema for the spatial characteristics of coverages. Coverages support mapping from a spatial, temporal, or spatiotemporal domain to feature attribute values where feature attribute types are common to all geographic positions within the domain. A coverage domain consists of a collection of direct positions in a coordinate space that may be defined in terms of up to three spatial dimensions, as well as a temporal dimension. This abstract standard was developed in collaboration with ISO/TC 211 and is identical to ISO 19123 *Coverages*.

Topic 7: Earth Imagery

This abstract specification [106] defines a reference model for standardization in the field of geographic imagery. This reference model identifies the scope of the standardization activity being undertaken and the context in which it takes place. The scope includes gridded data with an emphasis on imagery. Although structured in the context of information technology and information technology standards, this technical specification is independent of any application development method or technology implementation approach.

Topic 8: Relationships Between Features

Topic 5 of the abstract specification introduces features, an abstraction of the entities in the real world. Entities in the real world do not exist in isolation. Typically, an entity in the real world is related to other real-world entities in a variety of ways. This topic [107] introduces an abstraction for the relationships between entities in the real world.

Topic 10: Feature Collections

A feature collection is an abstract object consisting of feature instances, their feature schema, and project schema. This document [108] discusses the need (or not) for these concepts and how feature collections can be used.

Topic 11: Metadata

This topic [63] refers to ISO 19115-1:2014 *Metadata – Part 1: Fundamentals*, which defines a comprehensive metadata schema that is used to fully describe geographic resources. Minimum metadata may be used in resource catalogues and portals for discovery purposes, or comprehensive metadata can be used to support a complete understanding of resources, allowing them to be used properly to their full potential.

Topic 12: The OGC Service Architecture

This abstract standard [109] identifies and defines architecture patterns for service interfaces used for geographic information and to explain the relationship to the open systems environment model. It presents a geographic services taxonomy and a list of example geographic services placed in the services taxonomy. It also prescribes how to create a platform-neutral service specification and how to derive platform-specific service specifications that are conformant with this. The standard provides guidelines for the selection and specification of geographic services from both platform-neutral and platform-specific perspectives. It is identical to ISO 19119 *Services*.

Topic 13: Catalogue Services

This abstract standard [110] covers geospatial information access services, which include geospatial information retrieval services, geospatial product information services, and geospatial catalogue services. This topic, thus, covers OGC services for both data discovery and data access. It defines the term "catalogue" to describe the set of service interfaces that support organization, discovery, and access of geospatial information. Catalogue services help users or application software to find information that exists anywhere in a distributed computing environment. A catalogue can be thought of as a specialized database of information about geospatial resources available to a community of users. These resources are assumed to have the OGC feature, feature collection, catalogue, and metadata interfaces, or they may be geoprocessing services. Catalogues assist in the organization and management of diverse geospatial data and services for discovery and access, support discovery of resource information from diverse sources and gather it into a single, searchable location, and provide a means of locating, retrieving, and storing the resources indexed by the catalogue.

Topic 14: Semantics and Information Communities

This abstract specification [111] provides the essential model to permit interoperability across information communities. An information community is a collection of people (a government agency or group of agencies, a profession, a group of researchers in the same discipline, corporate partners cooperating on a project, etc.) who, at least part of the time, share a common digital geographic information language and share common spatial feature definitions; for example, if each information community has a fixed vocabulary, a fixed collection of schemas, and an unambiguous set of feature instances, a difference in vocabulary, say the use of "house" versus "dwelling", will inhibit interoperability. This abstract specification provides an essential model to overcome these differences and enable everyone to be able to share information with each other, as if they were native.

Topic 15: Image Exploitation Services

This topic volume [112] describes the categories and taxonomy of image exploitation services needed to support the use of images and certain related coverage types. Image exploitation services are required to support most aspects of image exploitation, including precision measurement of ground positions and of object dimensions; for example, a variety of services are needed for extracting features from images, or digital elevations from stereoscopic images. Image exploitation services are widely implemented and used in photogrammetric systems, currently using custom interfaces. Although the focus of this document is on services for using images, many of these services are also applicable to using other types of grid coverages and some nongrid coverages.

Topic 16: Image Coordinate Transformation Services

This topic [113] covers image coordinate conversion services; that is, this part of the abstract specification describes services for transforming image position coordinates to and from ground position coordinates. These services might alternatively be called *image geometry model services*.

Topic 17: Location-Based Services

This topic [114] covers location-based/mobile services; that is, this part of the abstract specification describes services that take advantage of mobility and the position or relative position of devices and points, lines, or polygons of service. Important concepts are location, route, and types of service.

Topic 18: Geospatial Digital Rights Management Reference Model

This abstract standard [115] is a reference model for digital rights management (DRM) functionality for geospatial resources (GeoDRM). As such, it is connected to the general DRM market in that geospatial resources must be treated as nearly as possible like other digital resources, such as music, text, or services.

Topic 19: Geographic Information – Linear Referencing

This abstract standard [116] specifies a conceptual schema for locations relative to a one-dimensional object as measurement along that object. It defines a description of the data and operations required to use and support linear referencing.

Topic 20: Observations and Measurements

This abstract standard [117] defines a conceptual schema for observations and for features involved in sampling when making observations. It provides models for the exchange of information describing observation acts and their results, both within and between different scientific and technical communities. Observations commonly involve sampling of an ultimate feature of interest. This abstract standard defines a common set of sampling feature types classified primarily by topological dimension, as well as samples for ex-situ observations. The schema includes relationships between sampling features (subsampling, derived samples).

Topic 21: Discrete Global Grid Systems

This abstract standard [118] supports the specification of discrete global grid system (DGGS) implementation standards. A DGGS is a spatial reference system that uses a hierarchical tessellation of cells to partition the globe. DGGSs enable combined analysis of very large, multisource, multiresolution, multidimensional, distributed geospatial data.

Topic 22: Core Tiling Conceptual and Logical Models for 2D Euclidean Space

This abstract specification [119] describes a general conceptual model for tiling space in any dimension. It also defines a logical model for the tessellation of 2D Euclidean space.

OGC Implementation Standards

OGC implementation standards [120] are primarily focused on geospatially enabling the World Wide Web; technical documents that specify data management, interfaces, and/or encodings. Table 15.14 is a list of OGC standards as of November 2021.

A synopsis of OGC standards is provided here.

Table 15.14 Listing of OGC standards

Listing Number	Title
1.	Simple Feature Access – Part 1: Common Architecture
2.	Simple Feature Access – Part 2: SQL Option
3.	Simple Features Implementation Specification for CORBA
4.	Simple Features Implementation Specification for OLE/COM
5.	Web Map Service Implementation Specification
6.	Styled Layer Descriptor
7.	Symbology Encoding Implementation Specification
8.	Web Map Context Implementation Specification
9.	OWS Context Conceptual Model
10.	OWS Context Atom Encoding Standard
11.	OWS Context GeoJSON Encoding Standard
12.	Web Map Tile Service (WMTS) Implementation Standard
13.	Web Feature Service 2.0 Interface Standard
14.	Filter Encoding 2.0 Encoding Standard
15.	Web Coverage Service Interface Standard
16.	Web Coverage Processing Service (WCPS) Language Interface Standard
17.	Web Service Common Implementation Specification
18.	Web Processing Service
19.	Publish/Subscribe Interface Standard 1.0 – Core
20.	Publish/Subscribe Interface Standard 1.0 – Soap Protocol Binding Extension
21.	Geography Markup Language (GML) Encoding Standard
22.	City Geography Markup Language (CityGML) Encoding Standard
23.	GeoScience Markup Language (GeoSciML)
24.	IndoorGML
25.	Land and Infrastructure Modeling Standards (InfraGML) Parts 0-7
26.	GML in JPEG 2000 for Geographic Imagery Encoding Specification
27.	Catalogue Service Specification
28.	CSW-ebRIM Registry Service – Part 1: ebRIM profile of CSW
29.	Catalogue Services Specification 2.0.2 – ISO Metadata Application Profile
30.	Catalogue Services Standard 2.0 Extension Package for ebRIM Application Profile: Earth Observation Products
31.	Catalogue Services 3.0 – General Model
32.	Catalogue Services 3.0 – HTTP Protocol Binding
33.	Ordering Services Framework for Earth Observation Products Interface Standard
34.	OpenSearch Geo and Time Extensions
35.	OpenSearch for Earth Observation
36.	KML
37.	Location Services (OpenLS) Implementation Specification
38.	Coordinate Transformation Service Implementation Specification
39.	GeoAPI Implementation Standard
40.	Geospatial eXtensible Access Control Markup Language (GeoXACML)
41.	Network Common Data Form (NetCDF) Standard
42.	Observations and Measurements
43.	TimeseriesML
44.	WaterML series of standards
	OGC WaterML 2.0: Part 1 – Timeseries
	OGC WaterML 2: Part 2 – Ratings, Gauges and Sections
	OGC WaterML 2: Part 3 – Surface Hydrology Features (HY_Features) – Conceptual Model
	OGC WaterML 2: Part 4 – GroundWaterML 2 (GWML2)
45.	SWE Common Data Model Encoding Standard
46.	SWE Service Model Implementation Standard
47.	Sensor Model Language (SensorML)
48.	Sensor Observation Service Interface Standard
49.	Sensor Planning Service Implementation Standard

Table 15.14 (Continued)

Listing Number	Title
50.	SensorThings API Part 1: Sensing
51.	Georeferenced Table Joining Service Implementation Standard
52.	Well known text representation of coordinate reference systems
53.	3-d Portrayal Service
54.	Augmented Reality Markup Language 2.0 (ARML 2.0)
55.	CDB Standard
56.	GeoPackage Encoding Standard
57.	GeoSPARQL – A Geographic Query Language for RDF Data
58.	GeoRSS Encoding Standard
59.	Open GeoSMS Standard
60.	Geospatial User Feedback (GUF) Standard
61.	Indexed 3-d Scene Layer (I3S) and Scene Layer Package Format Specification
62.	Moving Features Access
63.	Open Modeling Interface (OpenMI) Interface Standard
64.	3D Tiles Specification
65.	CityJSON Community Standard 1.0
66.	EO-GeoJSON
67.	GeoTIFF Standard
68.	Hierarchical Data Format Version 5 (HDF5®) Core Standard
69.	OGC API Common Part 1: Core
70.	OGC API Common Part 2: Geospatial Data (draft)
71.	API – Environmental Data Retrieval
72.	API – Features
73.	OGC API – Features Part 2: Coordinate Reference System by Reference
74.	PipelineML Conceptual and Encoding Model Standard
75.	OWS Security
76.	Semantic Sensor Network (SSN)
77.	Symbology Core
78.	Time ontology in OWL
79.	Two Dimensional Tile Matrix Set
80.	LAS

1. OGC Implementation Specification for Geographic Information – Simple Feature Access – Part 1: Common Architecture

Simple feature application programming interfaces (APIs) [121] provide for publishing, storage, access, and simple operations on simple features (points, lines, polygons, multipoints, etc.). The purpose of these specifications is to describe interfaces to allow GIS software engineers to develop applications that expose the functionality required to access and manipulate geospatial information comprising features with simple geometry using different technologies. This part of OGC Simple Features Access (SFA) describes the common architecture for simple feature geometry. It is identical to ISO 19125-1. The simple feature geometry object model is distributed computing platform neutral and is provided using UML notation. The base geometry class has subclasses for Point, Curve, Surface and GeometryCollection. Each geometric object is associated with a spatial reference system, which describes the coordinate space in which the geometric object is defined. The extended geometry model has specialized zero, one, and two-dimensional collection classes named MultiPoint, MultiLineString, and MultiPolygon for modeling geometries corresponding to collections of Points, LineStrings, and Polygons, respectively. MultiCurve and MultiSurface are introduced as abstract superclasses that generalize the collection interfaces to handle curves and surfaces. This part of OGC simple feature access implements a profile of the spatial schema described in ISO 19107:2019 *Geographic information – Spatial schema*. A detailed mapping of the schema in SFA to the schema described in ISO 19107:2019 is provided.

The three OGC simple features implementation specifications (one each for object linking and embedding component object model (OLE/COM), Common Object Broker Architecture (CORBA), and SQL) define interfaces that enable transparent access to geographic data held in heterogeneous processing systems on distributed computing platforms.

2. OGC Implementation Specification for Geographic Information Simple Feature Access – Part 2: SQL Option

This second part of OGC Simple Features Access (SFA) [122] defines a SQL schema that supports storage, retrieval, query, and update of feature collections via the SQL call-level interface (SQL/CLI) (ISO/IEC 9075-3:2003). A feature has both spatial and nonspatial attributes. Spatial attributes are geometry valued, and simple features are based on two or fewer dimensional geometric (point, curve, and surface) entities in two or three spatial dimensions with linear or planar interpolation between vertices. This standard is dependent on the common architectural components defined in Part 1 of this standard.

In an SQL implementation, a collection of features of a single type are stored as a *feature table*, usually with some geometry-valued attributes (columns). Each feature is primarily represented as a row in this feature table and described by that and other tables logically linked to this base feature table using standard SQL techniques. The nonspatial attributes of features are mapped onto columns whose types are drawn from the set of SQL data types, potentially including SQL3 user-defined types (UDT). The spatial attributes of features are mapped onto columns whose types are based on the geometric data types for SQL defined in this standard and its references. Feature-table schemas are described for two sorts of SQL implementations: implementations based on a more classical SQL relational model using only the SQL predefined data types and SQL with additional types for geometry. In any case, the geometric representations have a set of SQL-accessible routines to support geometric behavior and query.

In an implementation based on predefined data types, a geometry-valued column is implemented using a *geometry ID* reference into a geometry table. A geometry value is stored using one or more rows in a single geometry table, all of which have the geometry ID as part of their primary key. The geometry table may be implemented using standard SQL numeric types or SQL binary types; schemas for both are described in this standard.

The term "SQL with geometry types" is used to refer to a SQL implementation that has been extended with a set of geometry types. In this environment, a geometry-valued column is implemented as a column whose SQL type is drawn from this set of geometry types. The mechanism for extending the type system of an SQL implementation is through the definition of user-defined types. Commercial SQL implementations with user-defined type support have been available since mid-1997, and an ISO standard is available for UDT definition. This standard does not prescribe a particular UDT mechanism but specifies the behavior of the UDTs through a specification of interfaces that must be supported. These interfaces are described for SQL3 UDTs in ISO/IEC 13249-3.

3. OGC Simple Features Implementation Specification for CORBA

The purpose of this specification [123] is to provide interfaces to allow GIS software engineers to develop applications that expose functionality required to access and manipulate geospatial information comprising features with *simple* geometry using OMG's CORBA technology.

The Common Object Request Broker Architecture (CORBA) provides a specification for object-oriented distributed systems in a language, operating system, platform, and vendor-independent way. It has no ISO equivalent and has not been used much in recently.

4. OGC Simple Features Implementation Specification for OLE/COM

This specification [124] is based on use of the object linking and embedding, database (OLE DB) and ActiveX data objects (ADO) facilities for accessing data using Microsoft technology. As an OLE/component object model (COM)-based standard, current Microsoft technologies for database access are described with respect to geographic information processing. These technologies include open database connectivity (ODBC), data access object (DAO), remote data objects (RDO), ADO, and OLE DB. ADO specifically provides the OLE automation object-oriented standards for accessing and manipulating databases. This specification addresses the unique requirement of GIS-specific interfaces above and beyond the current interfaces available through current Microsoft data access technologies.

5. OGC Web Map Service (WMS) Implementation Specification

This standard [125] defines the behavior of a service that dynamically produces spatially referenced maps from geographic information. It specifies operations to retrieve a description of the maps offered by a server to retrieve a map and to query a server about features displayed on a map. It is applicable to pictorial renderings of maps in a graphical format; it is not applicable to retrieval of actual feature data or coverage data values. A WMS produces maps of spatially referenced data dynamically from geographic information. It defines a *map* to be a portrayal of geographic information as a digital image file suitable for display on a computer screen. A map is not the data itself. WMS-produced maps are generally rendered in a pictorial format, such as PNG, graphics interchange format (GIF), or JPEG, or occasionally as vector-based graphical elements in scalable vector graphics (SVG) or Web computer graphics metafile (WebCGM) formats. See Chap. 16 for more details.

6. OGC Styled Layer Descriptor (SLD) Profile of the WMS Implementation Specification
The styled layer descriptor (SLD) profile of the WMS encoding standard [126] defines an encoding that extends the WMS standard to allow user-defined symbolization and coloring of geographic feature and coverage data. SLD addresses the need for users and software to be able to control the visual portrayal of the geospatial data. The ability to define styling rules requires a styling language that both the client and server can understand. The OGC symbology encoding standard (SE) provides this language, while the SLD profile of WMS enables application of SE to WMS layers using extensions of WMS operations. Additionally, SLD defines an operation for standardized access to legend symbols. See Chap. 16 for more details.

7. OGC Symbology Encoding (SE) Implementation Specification
This specification [127] specifies the format of a map styling language for producing georeferenced maps with user-defined styling. This language can be used to portray the output of WMS, WFS, and Web coverage servers (WCS). It defines an XML encoding that can be used for styling feature and coverage data. These styles apply either to specific feature types or coverage types, depending on the used data type. Symbology encoding includes the FeatureTypeStyle and CoverageStyle root elements. These elements include all information for styling the data such as filter and different kinds of symbolizers. As symbology encoding is a grammar for styling map data independent of any service interface specification, it can be used flexibly by a number of services that style georeferenced information or store styling information that can be used by other services.

8. OGC Web Map Context Implementation Specification
This specification [128] applies to the creation and use of extensible markup language (XML)-encoded documents, which unambiguously describe the state, or context, of a WMS client application in a manner that is independent of a particular client and that might be utilized by different clients to recreate the application state. The specification is a companion specification to the OGC WMS implementation specification. WMS specifies how individual map servers describe and provide their map content. The present context specification states how a specific grouping of one or more maps from one or more map servers can be described in a portable, platform-independent format for storage in a repository or for transmission between clients. This description is known as a Web map context document, or simply a context. A context document includes information about the server(s) providing layer(s) in the overall map, the bounding box and map projection shared by all the maps, sufficient operational metadata for client software to reproduce the map, and ancillary metadata used to annotate or describe the maps and their provenance for the benefit of human viewers.

The specification contains an XML schema against which context XML can be validated. There are several possible uses for context documents:

- The context document can provide default startup views for particular classes of user. Such a document would have a long lifetime and public accessibility.
- The context document can save the state of a viewer client as the user navigates and modifies map layers.
- The context document can store not only the current settings but also additional information about each layer (e.g., available styles, formats, SRS, etc.) to avoid having to query the map server again once the user has selected a layer.
- The context document could be saved from one client session and transferred to a different client application to start up with the same context.
- Contexts could be catalogued and discovered, thus providing a level of granularity broader than individual layers.

OGC Web Services Context Document (OWS Context)
This is a collection of three standards, which describe the conceptual model, and several encodings fully describing a configured service set. It is an expansion and replacement for the Web Map Context Implementation Specification. It can be exchanged among clients and allows a set of configured information resources to be passed between applications primarily as a collection of OGC services such as Web Feature Services, Web Map Services, Web Map Tile Services, and/or Web Coverage Services. It provides a common operating picture in a manner that is independent of a particular client and that might be utilized by different clients to recreate the application state.

9. OGC OWS Context Conceptual Model
This standard [129] provides the requirements and conceptual model to be used by the context encoding standards to provide a context document that specifies a fully configured service set that can be exchanged between clients. It also provides use cases to provide:

1. Exchange of a common view or common operating picture for shared situational awareness
2. Exchange of discovery results from catalogue searches to avoid duplication of effort
3. Exchange of configuration and or results of an analysis or processing activity.

10. OGC OWS Context Atom Encoding Standard
This standard [130] describes the Atom encoding of the OWS Context conceptual model. It describes the mapping

of the classes and attributes in the conceptual model to the XML/Atom encoding for core and specific OGC services and data standards.

11. OGC OWS Context GeoJSON Encoding Standard
This standard [131] describes the GeoJSON encoding of the OWS Context conceptual model. It describes the mapping of the classes and attributes in the conceptual model to GeoJSON encoding for core and specific OGC services and data standards. GeoJSON [132] is a geospatial data interchange format based on JavaScript Object Notation (JSON).

12. OGC Web Map Tile Service (WMTS) Implementation Standard
This standard [133] provides digital maps using predefined image tiles. The service advertises the tiles it has available using a declaration in the service metadata. This declaration defines the tiles available in each layer in each graphical representation style, in each format, in each coordinate reference system, at each scale, and over each geographic fragment of the total area covered. The service metadata document also declares the communication protocols and encodings through which clients can interact with the server. Clients can use the service metadata document to request specific tiles. WMTS complements the existing Web Map Service standard. The WMS standard focuses on flexibility in the client request, enabling clients to obtain exactly the final image they want. A WMS client can request that the server creates a map by overlaying an arbitrary number of the map layers offered by the server, over an arbitrary geographic bound, with an arbitrary background color, at an arbitrary scale, and in any supported coordinate reference system. The client may also request that the map layers be rendered using a specific server-advertised style or even using a style provided by the client when the WMS server implements the OGC SLD standard. However, all this flexibility comes at a price: server image processing must deal with some number of connected clients, and there is only limited potential to cache images between the server and client, since most images are different.

As Web service clients have become more powerful, it has become possible to consider an alternative strategy, which forces the clients to perform image overlays themselves, and which limits the clients to requesting map images that are not at exactly the right position, thereby forcing the clients to mosaic the tiles obtained from the server and clip the set of tiles into a final image. This restriction of image requests to a fixed, predefined set allows for servers to be limited based on communication processing abilities rather than image processing abilities, because servers can prerender some or all of their images and can use image caching strategies. The fixed set of images also enables network providers to cache images between the client and the server, reducing latency and bandwidth use. Popular, nonstandardized, commercial implementations of this approach, such as Google Maps, Microsoft Virtual Earth, and Yahoo! Maps, have already shown that there are clear performance benefits to adopting this methodology. WMTS offers an approach to declaring the images that a client can request from a server, enabling a single type of client to be developed for all servers. The standard specifies WMTS in two stages. First, an abstract specification describes the semantics of the resources offered by the servers and requested by the client. This abstract definition specifies the semantics of the ServiceMetadata document, of the tile images or representations, and of the optional FeatureInfo documents providing descriptions of the maps at specific locations. Second, the standard specifies several different concrete exchange mechanisms between clients and servers in two different architectural styles. See Chap. 16 for more details.

13. OGC Web Feature Service 2.0 Interface Standard
This standard [134] defines how a Web service describes, queries/filters, and delivers geographic features. It is identical to ISO 19142:2010 *Geographic information – Web feature service*. It specifies a service that provides transactions on and access to geographic features in a manner independent of the underlying data store. Features are described and streamed to the client using GML. The standard supports:

- Discovery functions, which allow a service to be interrogated to determine its capabilities and to retrieve the application schema that defines the feature types that the service offers.
- Query functions, which allow features or values of feature properties to be retrieved from the underlying data store based upon constraints, defined by the client, on feature properties.
- Locking functions, which allow exclusive access to features for the purpose of modifying or deleting features.
- Transaction functions, which allow features to be created, changed, replaced, and deleted from the underlying data store.
- Stored query functions, which allow clients to create, drop, list, and describe parameterized query expressions that are stored by the server and can be repeatedly invoked using different parameter values. See Chap. 16 for more details.

14. OGC Filter Encoding (FE) 2.0 Encoding Standard
This standard [135] defines the encoding of expressions that support the querying/filtering of usually large data stores to produce a subset of data that contains just the desired information. It is identical to ISO 19143:2010 *Geographic information – Filter encoding*. It describes an XML and Key–Value pair (KVP) encoding of a system-neutral syntax for expressing projections, selection, and sorting clauses col-

lectively, called a query expression. These components are modular and are intended to be used together or individually by services described in other standards. The standard defines an abstract component, named AbstractQueryExpression, from which other specifications can subclass concrete query elements to implement query operations. The XML representation is easily validated, parsed, and transformed into a server-specific language required to retrieve or modify object instances stored in an object store, e.g., a large data store of geographic features.

The FE standard defines the XML encoding for the following predicates (sets of computational *operations* applied to a data instance which evaluate to true or false):

- A standard set of logical predicates: and, or, and not.
- A standard set of comparison predicates: equal to, not equal to, less than, less than or equal to, greater than, greater than or equal to, like, is null, and between.
- A standard set of spatial predicates: equal, disjoint, touches, within, overlaps, crosses, intersects, contains, within a specified distance, beyond a specified distance, and BBOX.
- A standard set of temporal predicates: after, before, begins, begun by, contains, during, ends, equals, meets, met by, overlaps, and overlapped by.
- A predicate to test whether the identifier of an object matches the specified value (Sect. 15.3.11).

15. OGC Web Coverage Service (WCS) Interface Standard
This suite of standards specifies services and data models for delivering multidimensional coverage data over the World Wide Web. Grid coverages provide digital geospatial information representing space-varying phenomena for regularly spaced locations along zero, one, two, or three axes of a spatial coordinate reference system. Coverages may also have a time dimension, which may be regularly or irregularly spaced. A coverage defines, at each location in the domain, a set of fields that may be scalar valued (such as elevation) or vector valued (such as wind speed and direction) for specific bands of the electromagnetic spectrum. These fields (and their values) are known as the range of the coverage. A WCS provides access to potentially detailed and rich sets of geospatial information, in forms that are useful for client-side rendering, multivalued coverages, and input into scientific models and other clients. Where the WMS provides static images, and the WFS provides streams of feature data, the Web coverage service: provides detailed descriptions of the data available from a server; defines a rich syntax for requests against these data; and returns coverage data which may be interpreted, extrapolated, etc., and not just portrayed.

The suite of standards consists of a core service interface standard with service extensions and a core coverage standard with data model and format extensions (Fig. 15.51):

- Service standards
  - OGC Web Coverage Service (WCS) Interface Standard – Core [137]
  - OGC Web Coverage Service (WCS) Interface Standard –Transaction Extension [138]
  - OGC Web Coverage Service (WCS) Interface Standard – Range Subsetting Extension [139]
  - OGC Web Coverage Service (WCS) Interface Standard – Scaling Extension [140]
  - OGC Web Coverage Service (WCS) Interface Standard – CRS Extension [141]
  - OGC Web Coverage Service (WCS) Interface Standard – Interposation Extension [142]
  - OGC Web Coverage Service (WCS) Interface Standard – Processing Extension [143].
- Protocol extensions:
  - OGC Web Coverage Service (WCS) Interface Standard – KVP Protocol Binding Extension [144]
  - OGC Web Coverage Service (WCS) Interface Standard – XML/POST Protocol Binding Extension [145]
  - OGC Web Coverage Service (WCS) Interface Standard – XML/SOAP Protocol Binding Extension [146].
- Data model standards and extensions:
  - OGC Coverage Implementation Schema (CIS) [147]
  - OGC Coverage Implementation Schema – ReferenceableGridCoverage Extension [148].
  - Coverage format extensions:
  - OGC GML Application Schema – Coverages – GeoTIFF Coverage Encoding Profile [149]
  - OGC GML Application Schema – Coverages JPEG2000 Coverage Encoding Extension [150]
  - CF-netCDF 3.0 encoding using GML Coverage Application Schema [151].

16. Web Coverage Processing Service (WCPS) Language Interface Standard
This standard [152] defines a protocol-independent language for the extraction, processing, and analysis of multidimensional gridded coverages representing sensor, image, or statistics data. Services implementing this language provide access to original or derived sets of geospatial coverage information, in forms that are useful for client-side rendering, input into scientific models, and other client applications. It provides the processing language used by Web Coverage Service Interface Standard – Processing Extension.

17. OGC Web Service Common Implementation Specification
This document [153] specifies many of the behaviors that may be common to many OGC Web service standards. The behaviors specified by OWS common currently include:

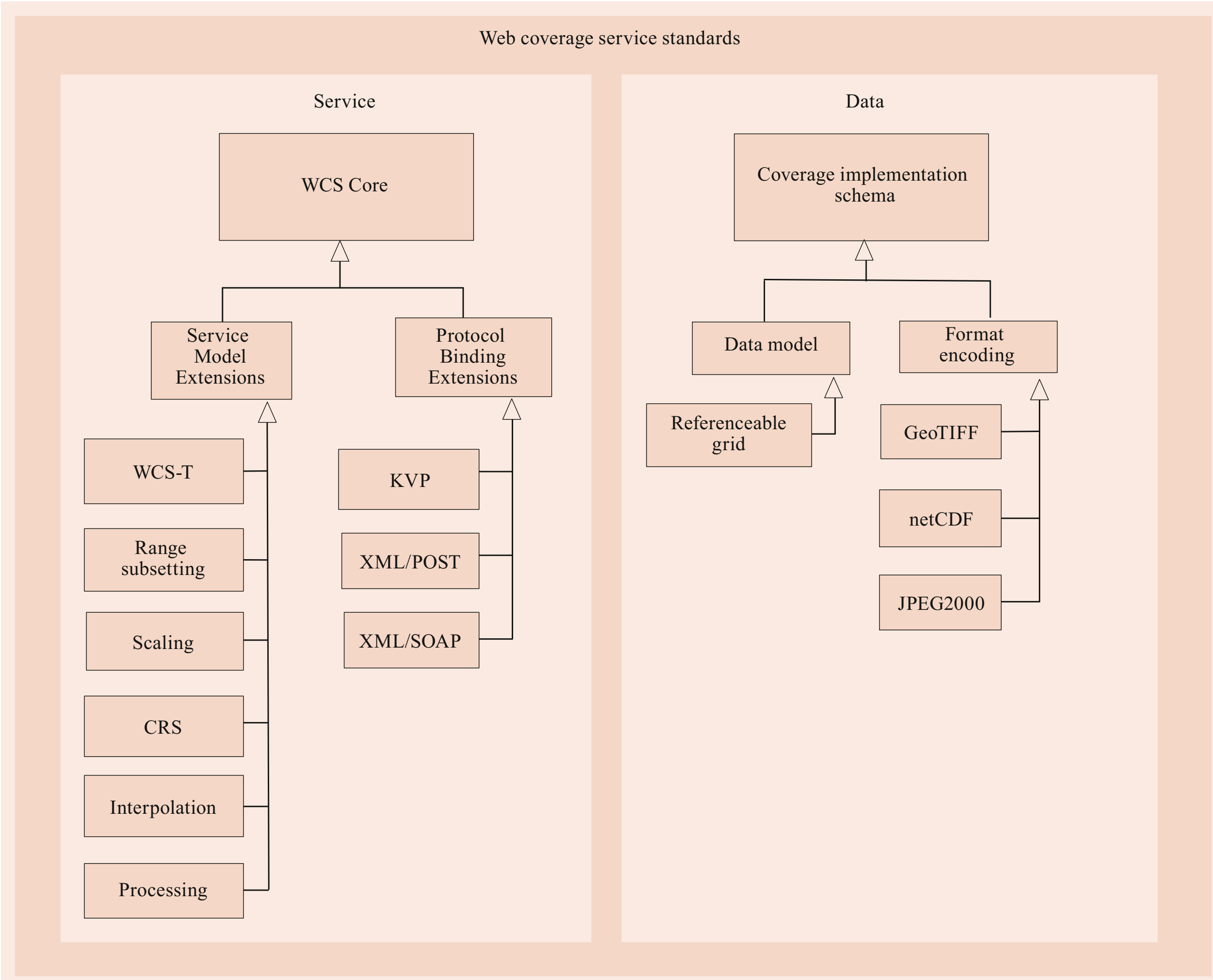


Fig. 15.51 Web Coverage Service standards. (After [136])

1. Operation request and response contents
2. Parameters and data structures included in operation requests and responses
3. XML and KVP encoding of operation requests and responses.

This standard is used to specify common behaviors used in many versions of OGC Web service standards. Rather than continuing to repeat this material, each specification will normatively reference relevant parts of OWS common. This document serves as a normative reference for present and future versions of OGC Web services; it is presently not used by WMS.

18. OGC Web Processing Service (WPS)
WPS [154] defines a standardized interface that facilitates the publishing of geospatial processes and the discovery of, and binding to, the discovered processes by clients. Processes include any algorithm, calculation, or model that operates on spatially referenced data. Publishing means making available machine-readable binding information, as well as human-readable metadata that allows service discovery and use. A WPS can be configured to offer any sort of GIS functionality to clients across a network, including access to preprogrammed calculations and/or computation models that operate on spatially referenced data. A WPS may offer calculations as simple as subtracting one set of spatially referenced numbers from another (e.g., determining the difference in influenza cases between two different seasons), or as complicated as a global climate change model. The data required by the WPS can be delivered across a network or available at the server. This interface specification provides mechanisms to identify the spatially referenced data required by the calculation, initiate the calculation, and manage the output from the calculation, so that the client can access it. This Web processing service is targeted at processing both vector and raster

data. The WPS specification is designed to allow a service provider to expose a Web-accessible process, such as polygon intersection, in a way that allows clients to input data and execute the process with no specialized knowledge of the underlying physical process interface or API. The WPS interface standardizes the way processes and their inputs/outputs are described, how a client can request the execution of a process, and how the output from a process is handled.

Publish/Subscribe Interface Standard

This set of standards specifies an interface that supports the Publish/Subscribe message exchange pattern for OGC Web services and addresses subscription management capabilities, such as creating a subscription, renewing a subscription, and unsubscribing. It also allows Publish/Subscribe services to advertise and describe supported message delivery protocols, such as SOAP messaging, Atom Syndication Format (Atom), and Advanced Messaging Queuing Protocol (AMQP). It is provided in two documents:

19. OGC Publish/Subscribe Interface Standard 1.0 – Core

This standard [155] specifies the core conceptual framework, functionality, and mechanisms for enabling the Publish/Subscribe messaging pattern with OGC Web services.

20. OGC Publish/Subscribe Interface Standard 1.0 – SOAP Protocol Binding Extension

This standard [156] utilizes the core concepts and mechanisms from the OGC Publish/Subscribe Interface Standard 1.0 – Core to provide the Publish/Subscribe messaging pattern with OGC Web services using the SOAP protocol.

21. OGC Geography Markup Language (GML) Encoding Standard

GML [157] is an XML grammar for expressing geographical features. GML serves as a modeling language for geographic systems, as well as an open interchange format for geographic transactions on the Internet. As with most XML-based grammars, there are two parts: the schema that describes the document and the instance document that contains the actual data. A GML document is described using a GML schema. This allows users and developers to describe generic geographic datasets that contain points, lines, and polygons. However, the developers of GML envision communities working to define community-specific application schemas that are specialized extensions of GML. Using application schemas, users can refer to roads, highways, and bridges instead of points, lines, and polygons. If everyone in a community agrees to use the same schemas, they can exchange data easily and be sure that a road is still a road when they view it. Clients and servers with interfaces that implement the WFS read and write GML data. GML is also an ISO standard (ISO 19136-1:2020) (Sect. 15.3.5).

22. OGC City Geography Markup Language (CityGML) Encoding Standard

This document [158] is an encoding standard for the representation, storage, and exchange of virtual 3-D city and landscape models. CityGML is implemented as an application schema of the GML version 3.1.1 (GML3). CityGML models both complex and georeferenced 3-D vector data along with the semantics associated with the data. In contrast to other 3-D vector formats, CityGML is based on a rich, general-purpose information model, in addition to geometry and appearance information. CityGML also provides an extension mechanism to enrich the data with identifiable features enhancing semantic interoperability for specific domains.

23. OGC Geoscience Markup Language 4.1 (GeoSciML)

This standard [159] defines a logical model and GML encoding rules for the exchange of geological features. It covers the broad domain of geology including earth materials, geological units and stratigraphy, geological time, geological structures, geomorphology, geochemistry and geologic sampling features. It also provides a simple version for the portrayal of geological features on digital maps.

24. OGC IndoorGML – with Corrigendum

This standard [160] provides a data model and an XML Schema for indoor navigation. It is an application schema of GML 3.2.1 for the representation and exchange of indoor navigation models. It supports the modeling of the topology and semantics of indoor spaces.

25. OGC Land and Infrastructure Modeling standards

These are a suite of standards that support the modeling of land and civil engineering infrastructure facilities and the land on which they are constructed and the exchange of that information using GML 3.2.1. The suite consists of a conceptual modeling standard: OGC Land and Infrastructure Conceptual Model Standard (LandInfra), [161] which specifies the vocabulary and UML conceptual models to provide the basic theory on which the multiple GML encoding standards are based. The encoding standards are provided in eight parts. OGC InfraGML 1.0: Part 0 – LandInfra Core – Encoding Standard [162] addresses the core requirements classes from LandInfra. The additional parts provide encoding for specific subject areas:

- OGC InfraGML 1.0 – Part 1 – LandInfra Land Features – Encoding Standard [163]
- OGC InfraGML 1.0 – Part 2 – LandInfra Facilities and Projects – Encoding Standard [164]
- OGC InfraGML 1.0 – Part 3 – LandInfra Alignments – Encoding Standard [165]
- OGC InfraGML 1.0 – Part 4 – LandInfra Roads – Encoding Standard [166]

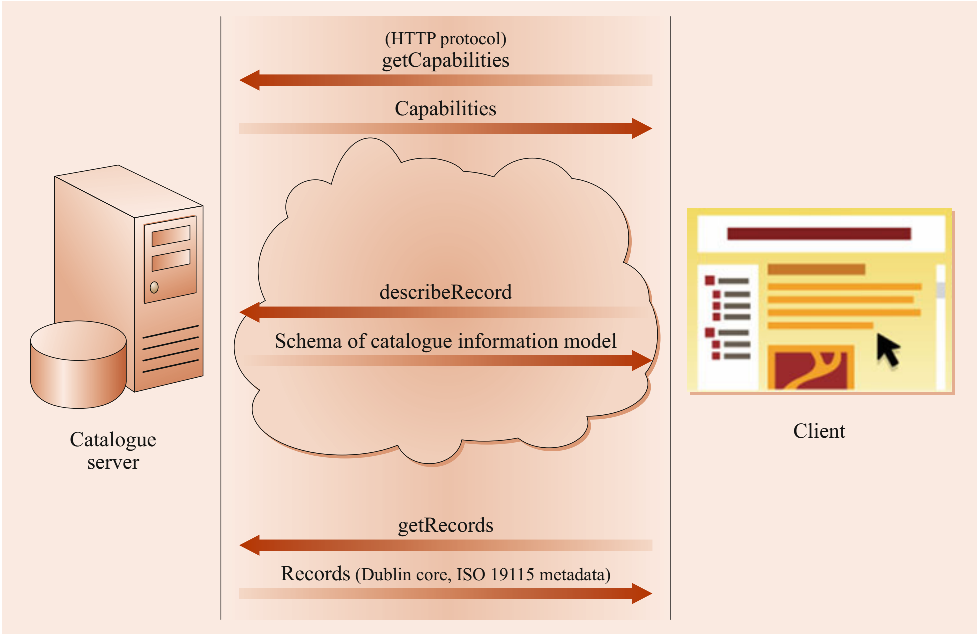


Fig. 15.52 Catalogue service Web application protocol operations

- OGC InfraGML 1.0 – Part 5 – LandInfra Railways – Encoding Standard [167]
- OGC InfraGML 1.0 – Part 6 – LandInfra Survey – Encoding Standard [168]
- OGC InfraGML 1.0 – Part 7 – LandInfra Land Division – Encoding Standard [169].

26. OGC GML in JPEG 2000 for Encoding Standard
This standard [170] defines the means by which GML is used within JPEG 2000 images for geographic imagery. The standard also provides packaging mechanisms for including GML within JPEG 2000 data files and specific GML application schemas to support the encoding of images within JPEG 2000 data files. JPEG 2000 is a wavelet-based image compression standard that provides the ability to include XML data for description of the image within the JPEG 2000 data file. This new version 2 takes advantage of the OGC Coverage Implementation Schema [147] as well as the ReferenceableGridCoverage Extension [148].

Catalogue Services
Catalogue services are used to support the discovery of, and access to, registered geographic resources. They support the ability to publish and search collections of metadata records for geographic resources: data, services, and related information. The metadata in the catalogues describes the resource's characteristics. The metadata in the catalogue can then be queried and presented for evaluation and processing by humans and software searching for resources. OGC has a multitude of catalogue service standards, extensions, and profiles. The two primary catalogue standards are presented below. OGC also has several standards that address the searching and ordering of geospatial and Earth-observation products and services, which are also covered here.

27.–30. OGC Catalogue Service Specification
This standard [171] specifies the interfaces, bindings, and a framework for defining application profiles required to publish and access digital catalogues of metadata for all types of geospatial resources: imagery, raster and vector data, services, studies, etc. Metadata provides the resource properties that can be queried and returned through catalogue services for evaluation and, in many cases, direct activation or retrieval of the referenced resource. Catalogue services support the use of one of several identified query languages to find and return results using well-known content models (metadata schemas) and encodings. The standard defines a general catalogue interface model and three application protocols: Z39.50 protocol binding, CORBA/IIOP (Internet Interobject) protocol binding, and Hypertext Transfer Protocol (HTTP) binding, referred to as catalogue services for the Web (CS-W) (Fig. 15.52). It also specifies the concept of harvesting, in which metadata records are harvested across the Web and pulled into a catalogue.

OGC maintains several application profiles of the catalogue services standard, each using a different information model: the base CSW using the Dublin Core information model; CSW-ebRIM Registry Service – Part 1: ebRIM profile of CSW [172]; OGC Catalogue Services Specification – ISO metadata application profile [173]; FGDC CSDGM application profile for CSW (a best practice); and OGC catalogue services standard extension package for ebRIM application profile: Earth observation products [174] (Fig. 15.53).

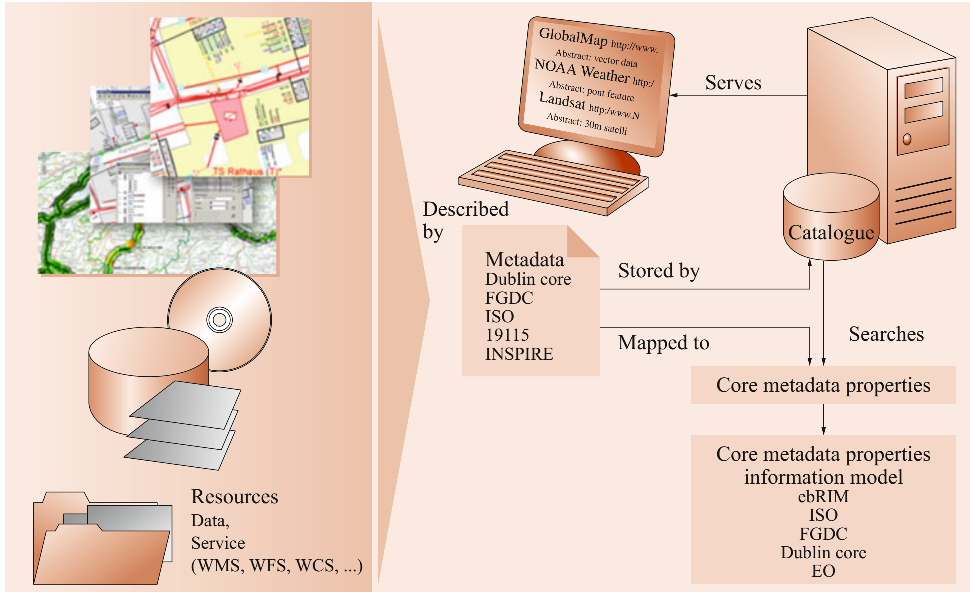


Fig. 15.53 CSW serving metadata based on different information models

31. OGC Catalogue Services 3.0 – General Model
This standard [175] is an update of OGC Catalogue Service Specification. It has been rewritten in the "modular core and extensions" style, with requirements and conformance classes clearly noted, as established by OGC for all their future standards. This document provides the abstract model for the common architecture specifying the abstract interfaces between clients and catalogue services. The architecture provided in this document is platform neutral. Specific implementation protocols, such as HTTP and OpenSearch, are provided in separate documents.

32. OGC Catalogue Services 3.0 Specification – HTTP Protocol Binding
This standard [176] specifies the HTTP implementation profile of the Catalogue Services – General Model also known at the Catalogue Services for the Web (CSW). It realizes the abstract interface of the general model using HTTP protocol bindings. HTTP protocol binding operation requests and responses are sent between clients and servers using HTTP GET and/or POST methods. It specifies three classes of service operations: OGC_Service, Discovery, and Manager:

- OGC_Service: allows the service to provide metadata about itself and a means of retrieving catalogue records based on their unique identifier.
- Discovery: facilitates the examination of services to retrieve runtime information about the data offered by the service, as well as a means of retrieving records from the catalogue using a general predicate language to define constraints that identify the subset of records to be retrieved.
- Manager: supports the addition, modification and removal of catalogue records:

Operation encodings use either a keyword-value pair (KVP) encoding for use with HTTP GET or an XML encoding for use with HTTP POST.

33. OGC Ordering Services for Earth Observation Products Interface Standard (OSEO)
This standard [177] specifies the interfaces, bindings, and encodings required to order Earth observation (EO) products in a heterogenous, distributed environment. It supports the ordering of EO products from previously identified dataset collections via standard catalogue interactions. It also provides a framework for XML extensions to enable the provision of EO products on media via email; via online protocols and online services, such as Web Coverage and Web Map services.

34. OGC OpenSearch Geo and Time Extensions
This standard [178] specifies extensions to OASIS OpenSearch [179] to provide a standard mechanism to query resources based on geographic extents or location names. It describes a simple format used to describe a search engine for use by search client applications and allows search engines to request specific query parameters from search clients. OpenSearch response elements can be used to extend syndication formats such as RSS and Atom.

35. OGC OpenSearch Extensions for Earth Observation
This standard [180] specifies extensions to OASIS OpenSearch [179] to provide a standard mechanism to query a repository of Earth observation information and allow for the syndication of repositories. It allows search engines to

inform clients about specific and contextual query parameters related to EO products, such as the originating platform or satellite, the sensor used to acquire the data, the processing center and software used to process the data, and satellite orbit information.

36. OGC KML

Google's KML (formerly Keyhole Markup Language) Vers. 2.2 was adopted as an OGC implementation standard in 2008. KML is an XML language focused on geographic visualization and the annotation of maps and images within in Internet-based two-dimensional maps and three-dimensional Earth browsers, such as KML viewer with drive, Google Earth, and ArcGIS Earth. KML provides the ability to present graphical data on a globe, add annotation to geographic data, specify icons and labels, define image overlays to attach to the ground or screen, define styles, write HTML descriptions or provide hyperlinks and embed images for KML features, and/or provide location and orientation of textured 3-D objects. Currently, KML 2.3 [181] utilizes certain geometry elements derived from GML 3.2. These elements include points, line strings, linear rings, and polygons.

37. OGC Location Service (OpenLS) Implementation Specification

This is a five-part standard [182] that specifies core interfaces that enable companies to provide location-based services (LBS) to mobile devices based on their location. Services include, e.g., emergency response (E-911, for example), personal navigators, traffic information services, proximity services, location recall, mobile field services, travel directions, restaurant finders, corporate asset locators, concierges, routing, vector map portrayals and interaction, friend finders, and geography voice-graphics (spoken directions), and many others. The five parts are: Part 1 – Directory Service, Part 2 – Gateway Service, Part 3 – Location Utility Service (geocoder/reverse geocoder service), Part 4 – Presentation (map portrayal) Service, and Part 5 – Route Service (Chap. 22).

38. OGC Coordinate Transformation Service Implementation Specification

This standard [183] provides a way for software to specify and access coordinate transformation services for use on specified spatial data. It addresses a key requirement for overlaying views of geodata (*maps*) from diverse sources, namely the ability to perform coordinate transformation in such a way that all spatial data are defined relative to the same spatial reference system.

39. OGC GeoAPI 3.0.1 Implementation Standard

This standard [184] provides a Java language application programming interface (API), which provides sets of types and methods for describing, managing, rendering, and manipulating geometric and geographic objects defined in ISO 19107, ISO 19111, and ISO 19115-1, and primitive types in ISO 19103. These are provided in the GeoAPI library, which is maintained by the GeoAPI project (http://www.geoapi.org). It is a realization of the old OGC Geographic Objects Implementation Specification, which it replaces.

40. Geospatial Extensible Access Control Markup Language (GeoXACML) Encoding Standard

This standard [185] defines a geospatial extension to the OASIS standard extensible access control markup language (XACML) [186] to provide geo-specific constraints. This extension incorporates spatial data types based on OGC simple features and GML to support authorization functions for the access or manipulation of map or geographic feature services based on spatial location. GeoXACML is a policy language that supports the declaration and enforcement of access rights for spatial jurisdictions such as states, provinces, or other restricted areas that can be used for access control for geospatial applications, such as those used in spatial data infrastructures.

41. OGC Network Common Data Form (NetCDF)

NetCDF is a suite of data model and encoding standards, application profiles, and extension mechanisms for geospatial information representing space and time-varying phenomena. NetCDF specifies a data model for array-oriented scientific data. It is commonly used for climate forecast (CF) data and can be used for a wide variety of multidimensional data. A freely distributed collection of access libraries implementing support for the data model and a machine-independent format are available. Together, the interfaces, libraries, and format support the creation, access, and sharing of multidimensional scientific data. Fig. 15.54 provides an overview of the netCDF suite of standards.

The suite of netCDF standards consists of (as of November 2021):

- OGC Network Common Data Form (NetCDF) Core Encoding Standard [187] specifies the core "classic" model and extension mechanisms.
- OGC CF-netD+CDF 3.0 encoding using GML Coverage Application Schema [151] required for implementing OGC Coverage Implementation Schema (CIS).
- NetCDF Binary Encoding Extension Standard: NetCDF Classic and 64-bit Offset Format (1.0) [188] specifies the netCDF classic (netCDF-3) and 64-bit offset file encoding formats.
- CF-netCDF3 Data Model Extension Standard (3.1) [189] specific the CF-netDCDF data model mapping onto the ISO 19123 Coverage schema and encoded using the NetCDF Binary Encoding Extension Standard described above.

15

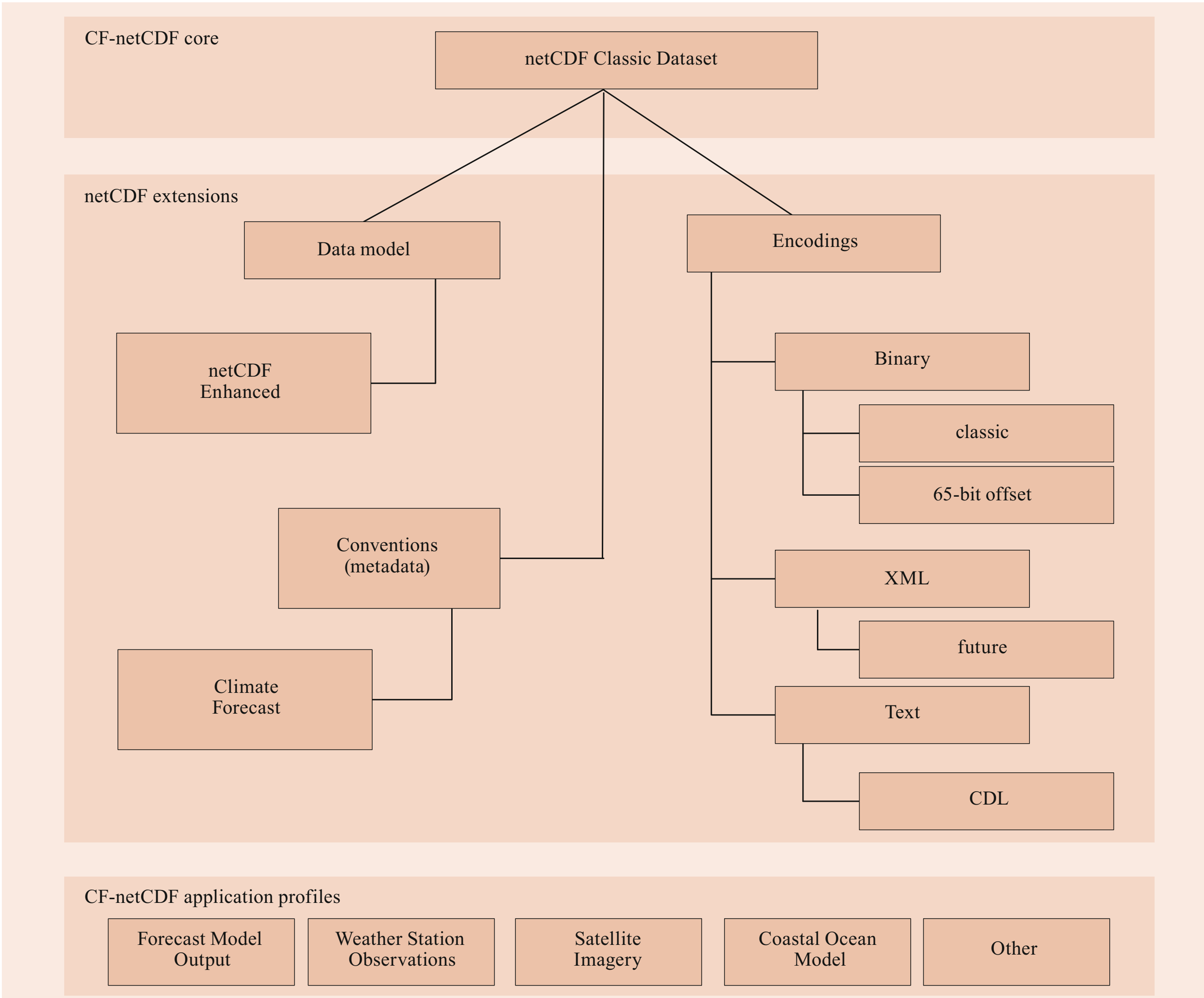


Fig. 15.54 OGC CF-netCDF standard hierarchy overview

- OGC Network Common Data Form (netCDF) NetCDF Enhanced Data Model Extension Standard (1.0) [190] specifies an enhanced data model used in netCDF-4.

42. Observations and Measurements – XML Implementation

This standard [191] specifies an XML implementation of the observations and measurements (O&M) conceptual model (OGC observations and measurements v2.0, also published as ISO 19156), including a schema for sampling features. This encoding is used by the OGC Sensor Observation Service (SOS) interface standard. More specifically, this standard defines XML schemas for observations, and features involved in sampling when making observations. These provide document models for the exchange of information describing observation acts and their results.

Timeseries Profile of Observations and Measurements
This standard [192] provides a conceptual model for a profile of observation and measurements for the representation of observations data as timeseries.

Earth Observation Metadata profile of Observation and Measurements
This standard [193] provides a schema for encoding metadata for describing and cataloguing products from sensors aboard Earth Observation satellites.

43. TimeseriesML 1.0 – XML Encoding of the Timeseries Profile of Observation and Measurements

This standard [194] defines an XML encoding of the Timeseries Profile of Observation and Measurements.

44. OGC WaterML 2.0

WaterML is a suite of standards and best practice papers providing conceptual information models and XML encodings for hydrologic observation data. The four standards are described below:

OGC WaterML 2.0: Part 1 – Timeseries

This standard [195] is an encoding standard for the representation of time series structures of hydrological observation data. WaterML 2.0 is an application schema of the Geography Markup Language 3.2.1 and makes use of the OGC Observation and Measurements suite of standards.

OGC WaterML 2: Part 2 – Ratings, Gauges and Sections

This standard [196] defines an information model to describe hydrological ratings, gauging observations, and survey observations for use in WaterML2.0.

OGC WaterML 2: Part 3 – Surface Hydrology Features (HY_Features) – Conceptual Model

This standard [197] defines a common conceptual feature model for hydrographic features independent of their geometric representation and scale. The model describes types of surface hydrologic features by defining fundamental relationships among various components of the hydrosphere. This includes relationships such as hierarchies of catchments, segmentation of rivers and lakes, and the hydrologically determined topological connectivity of features, such as catchments and waterbodies. It uses established models and patterns in use in the hydrology domain and endorsed by WMO and UNESCO, such as those documented in the "International Glossary of Hydrology". The standard also defines normative requirements for HY_Features implementation schemas and mappings to meet in order to be conformant with the conceptual model.

OGC WaterML 2: Part 4 – GroundWaterML 2 (GWML2)

This standard [198] specifies a conceptual and logical model and encoding standard for groundwater data: Ground Water ML (GWML2). GWML2 is an application schema of GML 3.2.1.

45. OGC SWE Common Data Model Encoding Standard

This standard [199] defines low-level data models used to define the representation, nature, structure and encoding of sensor data. These models allow applications and/or servers to structure, encode, and transmit self-describing and semantically enabled sensor datasets. This includes static as well as dynamically generated datasets. This standard applies to all categories of sensor types, from satellite imagery to simple in-situ sensors to full motion video.

46. OGC SWE Service Model Implementation Standard

This standard [200] specifies datatypes and interfaces common to all sensor Web services. It defines eight packages. The packages are: (1) Contents – defines data types that can be used in specific services that provide (access to) sensors; (2) Notification – defines the data types that support provision of metadata about the notification capabilities of a service as well as the definition and encoding of SWE service events; (3) Common – defines data types common to other packages; (4) Common codes – defines commonly used lists of codes with special semantics; (5) DescribeSensor – defines the request and response types of an operation used to retrieve metadata about a given sensor; (6) UpdateSensorDescription – defines the request and response types of an operation used to modify the description of a given sensor; (7) InsertSensor – defines the request and response types of an operation used to insert a new sensor instance at a service; and (8) DeleteSensor – defines the request and response types of an operation used to remove a sensor from a service.

47. OGC Sensor Model Language (SensorML)

This standard [201] specifies models and XML encoding that provide a framework within which the geometric, dynamic, and observational characteristics of sensors, sensor systems, and their metadata can be described. There are many different sensor types, from simple visual thermometers to complex Earth-observing satellites. These can all be supported through the definition of atomic process models and process chains. Within SensorML, all processes and components are encoded as application schema of the Feature model in the GML. This is one of the OGC SWE suites of standards.

48. OGC Sensor Observation Service (SOS)

The SOS [202] provides a Web service interface for managing deployed sensors and retrieving sensor descriptions, observations, and computational representations of observed features. It defines how to register new sensors and how to remove existing ones. It uses the observations and measurements – XML implementation standard. Interfaces are defined in a binding independent way. Key–Value Pair (KVP) and simple object access protocol (SOAP) bindings are specified in this standard. An SOS provides the ability to discover, bind to and interrogate individual sensors, sensor platforms, or networked constellations of sensors. This is one of the OGC SWE suites of standards.

49. OGC Sensor Planning Service Implementation Standard (SPS)

This standard [203] provides a standard interface to task sensors, satellites, and other information gathering assets. It defines interfaces for queries that provide information about the capabilities of a sensor and how to task the sensor. The standard is designed to support queries that have the following purposes: to determine the feasibility of a sensor planning request; to submit and reserve/commit such a request; to inquire about the status of such a request; to update or cancel such a request; and to request information about other OGC Web services that provide access to the data collected by the requested task. This is one of the OGC SWE suites of standards.

50. OGC SensorThings API Part 1: Sensing
This standard [204] provides a way to manage and retrieve observations and metadata from geospatial-enabled Internet of Things (IoT) devices, data, and applications over the Web. IoT is a network of physical devices (cars, appliances, heart monitors, etc.) connected over a network allowing them to exchange data and interact with or without human interaction. A Tasking, Part 2, is in development.

51. OGC Georeferenced Table Joining Service (TJS) Implementation Standard
This OGC standard [205] defines a simple way to describe and exchange tabular data that contains information about geographic objects. It standardizes an interface definition for and applies it to the creation and use of a table joining service (TJS). It includes the definition of all TJS requests and responses, including the specification of the Geographic Data Attribute Set (GDAS) encoding format using XML.

52. Geographic information – Well-Known Text Representation of Coordinate Reference Systems
This standard [206] defines the structure and content of a text string for describing coordinate reference systems. It is an implementation of the abstract model for coordinate reference systems described in ISO 19111. Coordinate reference systems are described in a compact machine and human readable form. It extends the original Well-known Text (WKT), originally described in ISO 19125, to allow for the description of coordinate operations. It is identical to ISO 19162.

53. OGC 3-D Portrayal Service
This standard [207] specifies a service interface for Web-based 3-D geographic data portrayal. It supports the delivery of 3-D scene data and server-side 3-D scene rendering. It is applicable to 3-D geographic data stores that target a range of portrayal clients and clients designed to portray 3-D data. It provides semantic interoperability, representing 3-D geodata with rich metadata and technical requirements for portrayal by a client. It does not define a transmission format for scenes or images.

54. OGC Augmented Reality Markup Language (ARML)
This standard [208] specifies an interchange format using XML grammar: Augmented Reality Markup Language (ARML), which lets users describe virtual objects in an augmented reality (AR) scene. It is used to describe virtual objects placed into an AR environment and their registration relative to the real world. Devices running an AR implementation, using ARML, use a camera along with orientation and GPS sensors and a screen for viewing the real world onto which the virtual objects are projected.

55. CDB
This suite of standards is designed to be used in the modeling and simulation discipline to provide a common setting in which interconnected simulators share a common view of a synthetic environment. This synthetic environment can represent activities and Earth features that can provide realistic simulations of the real world, such as warfighting in theaters of war or manufacturing processes in factories. It provides a model and structure for a data store of a virtual representation of the Earth that can be used by a single computer or a large distributed network of systems. This standard was originally developed in the US defense community as the common data base. The open format for the storage, access, and modification of a synthetic environment database provides an online repository where simulator computers, servers, and clients can simultaneously retrieve and modify data in real time allowing networked simulators to interact. Presently, it consists of 14 documents, a mixture of OGC Best Practices and OGC Standards (Volume 0 through 12 plus one extension). For an overview of the CDB, see Volume 0: Primer for the OGC CDB Standard: Model and Physical Data Store Structure [209]. The OGC standards are:

- Volume 1: OGC CDB Core Standard: Model and Physical Data Store Structure [210]
- Volume 3: OGC CDB Terms and Definitions [211]
- CDB Multi-Spectral Imagery Extension [212].

56. OGC Geopackage Encoding Standard
This standard [213] is an encoding standard designed for the exchange and direct use of vector geospatial features, and/or tile matrix sets of images and raster maps, and metadata tables using SQLite (http://www.sqlite.org) extensions. It defines an SQL database schema designed for use with the SQLite software library. Geopackages are especially useful on cell phones, tablets, and other mobile devices, and are designed to work in limited network or disconnected environments.

57. OGC GeoSPARQL – A Geographic Query Language for RDF Data
This standard [214] defines a set of rules and a vocabulary for representing geospatial data in a Resource Description

Framework (RDF), OWL Web Ontology Language, and SPARQL (a W3C RDF query language) to support representing and querying geospatial data on W3C's semantic web.

58. OGC GeoRSS Encoding Standard

This community standard [214, 215] specifies a way to extend really simple syndication (RSS) feeds with simple geographic location information. An RSS feed is a simple file that is submitted to a feed directory allowing the submitter/subscriber to see content (a Web page, for example) a short time after it is updated. It enables RSS and Atom applications to request, aggregate, share, and map geographically tagged feeds. There are two encodings of GeoRSS:

- GeoRSS-Simple: a very simple format for basic geometries (point, line, box, and polygon) in WGS-84 latitude/longitude
- GeoRSS GML for a greater range of features and additional coordinate reference systems.

59. OGC GeoSMS Standard – Core

This standard [216] specifies an extended format for providing simple location information content using the Short Message Service (SMS). SMS is the text communication service component for phone, Web, and mobile communication systems.

60. OGC Geospatial User Feedback Standard

This set of standards allows users of geospatial data products to provide feedback metadata, such as ratings, comments, quality reports, usage reports, citations of related datasets or publications describing usage, additional provenance information, and significant events based on their experiences with using the data. It is designed to be used by metadata catalogue servers and clients to manage and exchange user feedback information. It consists of two standards:

- OGC Geospatial User Feedback Standard: Conceptual Model [217] provides a UML data model specifying the classes and schema extending the ISO 19115-1 Metadata Fundamentals model.
- OGC Geospatial User Feedback Standard: XML encoding extension [218] provides an XML encoding based on the conceptual model compatible with ISO/TS 19115-3 Geographic Information – Metadata – Part 3: XML schema implementation for fundamental concepts.

61. OGC Indexed 3-d Scene Layer (I3S) and Scene Layer Package Format Specification

This community standard [219] provides a delivery format and persistence model for scene layers, referred to as Indexed 3-d scene layer (I3S) and scene layer packages (SLPK), designed to stream large amounts of heterogeneously distributed 3-D geospatial data while optimizing performance and scalability. An I3S dataset (a scene layer) is a container for large amounts of 3-D geographic data. The following layer types are specified in the standard:

- 3-D Objects (e.g., building exteriors from geospatial data and 3-D models)
- Integrated meshes (e.g., a mesh surface with high-resolution imager textures representing the skin of the Earth, typically created from satellite, aerial, or drone imagery)
- Point features (such as geolocated hospitals or schools, trees, street furniture, and signs).

I3S uses JSON and binary ArrayBuffers to support the 3-D content in a variety of coordinate reference systems and height models. I3S is designed to work with data stored locally, in the cloud, on the Web, and on mobile devices.

62. OGC Moving Features

OGC moving features consists of a set of three standards that provide encoding and access methods for the implementation of ISO 19141:2008 Geographic Information – schema for moving features [90], which specifies a method for describing the geometry of a feature that moves as a rigid body. The number of features encoded using these standards can be very large, many thousands of features, and can be described using common space–time coordinates:

- OGC Moving Features Encoding Part 1: XML Core [220] provides an XML and GML encoding for ISO 19141.
- OGC moving features access [221] provides implementation methods, based on abstract methods defined in ISO 19141, for accessing moving feature data. This includes database access to feature attributes, relationship information between a trajectory object and one or more geometry objects and/or between two trajectory objects.
- OGC moving features encoding extension: simple comma separated values (CVS) [222] provides a simple CVS style encoding for ISO 19141 applicable to handling very large amounts of moving features.

63. OGC Open Modeling Interface (OpenMI) Interface Standard

This standard [223] defines a way for independently developed computer simulations, environmental processes, and other models to exchange data during runtime. It is designed for the exchange of data between process models and other modeling tools, such as databases and analytical and visualization applications. The standard is designed to be implemented through the use of software tools for creating and running OpenMI compliant components provided by the OpenMI Association (www.openmi.org).

64. 3D Tiles Specification

3D Tiles [224] is an OGC Community Standard adopted by OGC; a direct copy of the 3D Tiles specification by CESIUM GS, Inc. It is designed for streaming and rendering massive 3D geospatial content such as Photogrammetry, 3D Buildings, BIM/CAD, Instanced Features, and Point Clouds. It defines a hierarchical data structure and a set of tile formats which deliver renderable content. 3D Tiles does not define explicit rules for visualization of the content; a client may visualize 3D Tiles data however it sees fit.

65. CityJSON Community Standard 1.0

CityJSON [225] is an OGC Community Standard specifying a JSON-based encoding for a well-documented subset of the CityGML data model (version 2.0.0). CityJSON defines how to store digital 3D models of cities and landscapes. The aim of CityJSON is to offer an alternative to the GML encoding of CityGML, which can be verbose and complex to read and manipulate. CityJSON aims at being easy-to-use, both for reading datasets and for creating them. It was designed with programmers in mind, so that tools and APIs supporting it can be quickly built.

66. EO-GeoJSON

This standard [226] defines an encoding for Earth Observation metadata for datasets using GeoJSON and JSON-LD. It provides a JSON encoding metadata based on the Earth Observation Metadata Profile of Observation and Measurements [193]. Using JSON-LD (linked data) allows each property to be explicitly defined as a URI. Using JSON Schema allows validation of instances. The encoded metadata can be used in a Service Oriented Architecture as well as a Resource Oriented Architecture.

67. GeoTIFF Standard

The GeoTIFF format was initially developed during the early 1990's by Niles Ritter and Mike Ruth. The objective was to leverage a mature platform independent file format (TIFF) by adding metadata required for describing and using geographic image data. TIFF met the requirements for an underlying format, as it was – and still is – lossless and extensible.

This standard [227] defines the Geographic Tagged Image File Format (GeoTIFF) by specifying requirements and encoding rules for using the Tagged Image File Format (TIFF) for the exchange of georeferenced or geocoded imagery. The GeoTIFF 1.1 standard formalizes the existing community GeoTIFF specification version 1.0 and aligns it with the addition of data to the EPSG Geodetic Parameter Dataset.

68. Hierarchical Data Format Version 5 (HDF5®) Core Standard

OGC Hierarchical Data Format Version 5 [228] supports the management of large and complex data. It provides a data model, a programming interface, and a storage model; it supports an unlimited variety of data types. The data model is presented as an encoding standard suitable for scientific and engineering applications that employ multidimensional numeric arrays to describe temporally and spatially varying phenomena. It is self-describing through user-defined metadata. It supports complex data relationships and dependencies using grouping and linking mechanisms.

69. OGC API Common Part 1: Core (Draft)

OGC is developing a suite of Application Programming Interface (API) standards to allow geospatial services and products to communicate using documented interfaces. OpenAPI (https://www.openapis.org) is used to define these reusable modular API building blocks. They use a resource oriented architecture that conforms to the principles of Representational State Transfer (REST). OGC Common is the core standard that provides the common foundation on which all OGC APIs are built. This draft standard [229] defines the offerings and capabilities of an OGC API and is made up of three elements:

- GET /
  Provides clients with a landing page as a starting point for using the API. A landing page can provide title, description and attribution as metadata elements to inform the client. As clients follow links from the landing page they will be provided more resource specific metadata.
- GET /api
  Retrieves the API definition which describes the capabilities provided by that API
- GET /conformance
  Provides a list of modules that are implemented by the API.

70. OGC API Common Part 2: Geospatial Data (Draft)

This core part of the standard [230] provides a connection between the API landing page and resource-specific details:

- GET /collections
  Retrieves information which describes the set of supported Collections
- GET /collections/{collectionId}
  Retrieves descriptive information about a specific Collection
- GET /collections/{collectionId}items
  Retrieves the resources offered by a specific Collection

71. OGC API – EDR

OGC API – Environmental Data Retrieval Standard [231] provides a set of query patterns that provide access to temporal/geospatial data. The EDR API provides access to large complex collections of environmental data through a simple

Web interface. Data can be queried by Position, within an Area or a Cube, along a Trajectory, or through a Corridor.

72. OGC API – Features Part 1: Core

The OGC API – Features standard is composed of multiple parts, this part, Core [232], specifies discovery, query, and fetching operations that are restricted to features using the WGS 84 coordinate reference system. This standard provides an approach for interacting with fine grain access at the feature level rather than sharing data an entire dataset. This standard is identical to ISO 19168-1:2020.

73. OGC API – Features Part 2: Coordinate Reference System by Reference

This standard [233] extends Part 1 to specify how a server can declare the coordinate reference system used to present features, advertise a list of supported CRS identifiers, and how features can be accessed by geometry values or by using a bounding box in the supported CRSs. The list of supported reference systems CRS identifiers must be referenced by a uniform resource identifier (URI).

74. PipelineML Conceptual and Encoding Model Standard

The PipelineML Conceptual and Encoding Model Standard [234] defines concepts supporting the interoperable interchange of data pertaining to oil and gas pipeline systems. PipelineML supports the common exchange of oil and gas pipeline information.

75. OWS Security

This standard [235] specifies how OGC Web Services (OWS) advertise their Information Assurance (IA) Controls. It applies to OWS using HTTPS. It specifies how to identify and document IA controls for supporting authentication in a register and how a service can advertise their IA controls through existing capabilities documents.

76. Semantic Sensor Network (SSN)

This standard [236] provides an ontology for describing sensors. It provides a means of integrating OGC Sensor Web technologies with W3C Semantic Web technologies and Linked Data. It provides an ontology for describing sensors and their observations, the involved procedures, the observed feature of interest, the samples, and the observed properties and actuators. This standard is also a W3C Recommendation.

77. Symbology Core

OGC Symbology Conceptual Model: Core Part [237] provides a conceptual model for the portrayal of geographical data. The model consists of a minimal set of abstract classes representing extension points. It does not define any extensions. It is designed to be flexible to provide cartographic styling to fill the needs of a variety of information communities (weather symbols, thematic maps, nautical charts, etc.) independent of encoding.

78. Time Ontology in OWL

This standard [238] provides an ontology for describing the temporal properties of resources in the world or described in Web pages. It provides a vocabulary for expressing facts about topological relations among instances and intervals, information about durations and temporal positions, including date-time information. Time positions and durations may be expressed using the conventional calendar and clock or other reference systems such as geologic time or different calendars. This standard is also a W3C Recommendation.

79. Two Dimensional Tile Matrix Set

This standard [239] specifies the concepts of a Tile Matrix Set as a way of indexing space, based on a set of regular grids defining a domain (tile matrix) for a limited list of scales. Each tile matrix is divided into regular tiles distinctly identified by a tile column, a tile row, and a tile matrix identifier. It provides a data structure defining tile matrix properties and a data structure for defining a subset of a tile matrix set called a tile matrix set limits. It also provides XML and JSON encoding methods for both tile matrix sets and tile matrix set limits. It also contains a library of proposed tile matrix set definitions for a limited set of coordinate reference systems.

80. LAS

LAS Specification 1.4 [240] is an OGC Community Standard adopted by OGC originally published by the American Society for Photogrammetry and Remote Sensing (ASPRS). It specifies a format for LIDAR and other point cloud data records.

References

1. GSDI: SDI cookbook (2018). http://gsdiassociation.org/index.php/publications/sdi-cookbooks.html, Accessed 11 Nov 2018
2. Compvoets, J., Bregt, A., Rajabifard, A., Williamson, I.: Assessing the worldwide developments of national spatial data clearinghouses. Int. J. Geogr. Inf. Sci. **18**(7), 665–689 (2004)
3. Hjelmager, J., Delgado, T., Moellering, H., Cooper, A., Danko, D., Huet, M., Aalders, H., Martynenko, A.: Developing a modeling for the spatial data infrastructure (2007). https://icaci.org/files/documents/ICC_proceedings/ICC2005/htm/pdf/oral/TEMA6/Session%201/JAN%20HJELMAGER.pdf, Accessed 11 Nov 2018. XXIII Int. Cartogr. Conf., Moscow
4. US Federal Government: Federal Geographic Data Committee Washington, DC, US. http://www.fgdc.gov/nsdi/nsdi.html, Accessed 11 Nov 2018
5. Instituto Panamericano de Geografía e Historia: Organismo Especializado de la Organización de los Estados Americanos – OEA. http://ipgh.org, Accessed 11 Nov 2018
6. European Commission: INSPIRE, Brussels. http://inspire.jrc.ec.europa.eu/, Accessed 11 Nov 2018

7. Global Spatial Data Infrastructure Association (GSDI): Home page. http://www.gsdi.org/, Accessed 11 Nov 2018
8. University Consortium for Geographic Information Science (UCGIS), Alexanderia, VA, US: Home page. http://www.ucgis.org/, Accessed 11 Nov 2018
9. VIM: International Vocabulary of Basic and General Terms in Metrology. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, (1993)
10. US Department of Defense: Standard Practice: Mapping, Charting and Geodesy Accuracy, MIL-STD-600001. US Department of Defense, Washington D.C. (1990)
11. ISO 19157:2013: Geographic Information – Data Quality (2013)
12. Ponzio Jr., F.J.: Authoritative data source (ADS) framework and ADS maturity model 1. In: Proc. Ninth Int. Conf. Inf. Qual. ICIQ-04. (2004)
13. Diamond, J.: The curse of QWERTY (1997). http://discovermagazine.com/1997/apr/thecurseofqwerty1099/, Accessed 11 Nov 2018. Discover, April
14. The Oxford English Dictionary. Oxford University Press, Oxford (1970)
15. ISO/TC 211: Home page. https://committee.iso.org/home/tc211, Accessed 11 Nov 2018
16. ISO 19109:2015: Geographic information – rules for application schema (2015)
17. ISO 19110:2016: Geographic information – methodology for feature cataloguing (2016)
18. ISO 19101-1:2014: Geographic information – reference model – part 1: fundamentals (2014)
19. ISO/TS 19101-2:2018: Geographic information – reference model – part 2: imagery (2018)
20. ISO/TS 19103:2015: Geographic information – conceptual schema language (2015)
21. OMG: UML Resource Page. Object Management Group, Needham (2003)
22. ISO/IEC 19501:2005: Information technology – open distributed processing – Unified Modeling Language (UML) version 1.4.2 (2005)
23. ISO/IEC 10303-11:2004: Industrial automation systems and integration – product data representation and exchange – part 11: description methods: the EXPRESS language reference manual (2004)
24. ISO 19106:2004: Geographic information – profiles (2004)
25. ISO 19105:2000: Geographic information – conformance and testing (2000)
26. ISO/IEC 9646-1:1994: Information technology – open systems interconnection – conformance testing methodology and framework – part 1: general concepts (1994)
27. ISO/IEC 9646-2:1994: Information technology – open systems interconnection – conformance testing methodology and framework – part 2: abstract test suite specification (1994)
28. ISO/IEC 9646-5:1994: Information technology – open systems interconnection – conformance testing methodology and framework – part 5: requirements on test laboratories and clients for the conformance assessment process (1994)
29. ISO 19118:2011: Geographic information – encoding (2011)
30. ISO 19119:2016: Geographic information – services (2016)
31. ISO/IEC 10746-2:2009: Information technology – open distributed processing – reference model: foundations (2009)
32. ISO 19116:2019: Geographic information – positioning services (2019)
33. ISO 19111:2019: Geographic information – Spatial referencing by coordinates (2019)
34. ISO 19111-2:2009: Geographic information – spatial referencing by coordinates – part 2: extension for parametric values (2009)
35. Kresse, W., Fadaie, K.: ISO Standards for Geographic Information. Springer, Berlin, Heidelberg (2004)
36. ISO/TS 19127:2019: Geographic information – geodetic codes and parameters (2019)
37. ISO/CD 19161-1:2020: Geographic information – geodetic references – part 1: the international terrestrial reference system (ITRS) (2020)
38. ISO 19162:2019: Geographic information – well-known text representation of coordinate reference systems (2019)
39. ISO 6709:2008: Geographic information – standard representation of geographic point location by coordinates (2008)
40. ISO 19145:2013: Geographic information – registry of representations of geographic point location (2013)
41. ISO 19112:2019: Geographic information – spatial referencing by geographic identifiers (2019)
42. ISO 19155:2012: Geographic information – place identifier (PI) architecture (2012)
43. ISO 19155-2:2017: Geographic information – place identifier (PI) architecture – part 2: place identifier (PI) linking (2017)
44. ISO 19170-1:2021 Geographic information – Discrete Global Grid Systems Specifications – Part 1: Core Reference System and Operations, and Equal Area Earth Reference System
45. ISO 19107:2019: Geographic information – spatial schema (2019)
46. Jech, T.: Set theory. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Stanford University, Stanford (2002). available online at http://plato.stanford.edu/archives/fall2002/entries/set-theory/
47. ISO 8601-1:2019: Date and Time – Representations for information interchange – Part 1: Basic rules (2019)
48. OGC: The OpenGIS Abstract Specification, Topic 6: The Coverage Type and its Subtypes, Document 00-106. Open Geospatial Consortium, Wayland (2000)
49. W3C: SMIL: Synchronized Multimedia Integration Language (SMIL 2.1). W3C, Cambridge (2005). W3C Recommendation, 2005 12 13
50. ISO 19136-2:2015: Geographic information – Geography Markup Language (GML) – part 2: extended schemas and encoding rules (2015)
51. ISO 19137:2007: Geographic information – core profile of the spatial schema (2007)
52. ISO 19125-1:2004: Geographic information – simple feature access – part 1: common architecture (2004)
53. ISO 19123:2005: Geographic information – schema for coverage geometry and functions (2005)
54. ISO/TS 19130-1:2018: Geographic information – imagery sensor models for geopositioning – part 1: fundamentals (2018)
55. ISO/TS 19130-2:2014: Geographic information – imagery sensor models for geopositioning – part 2: SAR, InSAR, lidar and sonar (2014)
56. ISO/TS 19159-1:2014: Geographic information – calibration and validation of remote sensing imagery sensors and data – part 1: optical sensors (2014)
57. ISO/TS 19159-2:2016: Geographic information – calibration and validation of remote sensing imagery sensors and data – part 2: lidar (2016)
58. ISO/TS 19159-3:2018: Geographic information – calibration and validation of remote sensing imagery sensors and data – Part 3: SAR/InSAR (2018)
59. ISO/TS 19163-1:2016: Geographic information – content components and encoding rules for imagery and gridded data – part 1: content model (2016)
60. ISO 19117:2012: Geographic information – portrayal (2012)
61. ISO/TS 19104:2016: Geographic information – terminology (2016)

62. ISO 19146:2018: Geographic information – cross-domain vocabularies (2018)
63. ISO 19115-1:2014: Geographic information – metadata, part 1: fundamentals (2014)
64. ISO 15836:2017: Information and documentation – the Dublin core metadata element set – part 1: core elements (2017)
65. ISO 19115-2:2019: Geographic information – metadata – part 2: extensions for acquisition and processing (2019)
66. ISO/TS 19158:2012: Geographic information – quality assurance of data supply (2012)
67. ISO 19108:2002: Geographic information – temporal schema (2002)
68. ISO 19135-1:2015: Geographic information – procedures for item registration – part 1: fundamentals (2015)
69. ISO/TS 19135-2:2012: Geographic information – procedures for item registration – part 2: XML schema implementation (2012)
70. ISO 19126:2021: Geographic information – feature concept dictionaries and registers (2021)
71. ISO 19128:2005: Geographic information – web map server interface (2005)
72. De La Beaujardière, J. (ed.): Web Map Service Implementation Specification 1.1.1. OGC Document 01-068r3. Open Geospatial Consortium, Wayland (2002)
73. ISO 19142:2010: Geographic information – web feature service (2010)
74. ISO 19143:2010: Geographic information – filter encoding (2010)
75. ISO 19168-1:2020: Geographic information – geospatial API for features – part 1: core (2020)
76. ISO 19156:2011: Geographic information – observations and measurements (2011)
77. ISO: ISO/IEC Guide 99:2007, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM) (2007)
78. ISO 19153:2014: Geographic information – Geospatial Digital Rights Management Reference Model (GeoDRM RM) (2014)
79. ISO/TS 19150-1:2012: Geographic information – ontology – part 1: framework 1 (2012)
80. ISO 19154:2014: Geographic information – ubiquitous public access – reference model (2014)
81. ISO 14721:2012: Space data and information transfer systems – open archival information system (OAIS) – reference model (2012)
82. ISO 19165-1:2018: Geographic information – preservation of digital data and metadata – part 1: fundamentals (2018)
83. ISO 19132:2007: Geographic information – location-based services – reference model (2007)
84. ISO 19133:2005: Geographic information – location based services – tracking and navigation (2005)
85. Dijkstra, E.: A note on two problems in connection with graphs. In: Numerische Mathematik, vol. 1, pp. 269–271. Springer, Berlin, Heidelberg (1959)
86. Bellman, R.E.: On a routing problem. Q. Appl. Math. **16**, 87–90 (1958)
87. Ford Jr., L.R.: Network Flow Theory. Paper P-923. RAND, Santa Monica (1956)
88. ISO 19134:2007: Geographic information – location-based services – multimodal routing and navigation (2007)
89. ISO 19148:2021: Geographic information – geographic information – linear referencing (2021)
90. ISO 19141:2008: Geographic information – schema for moving features (2008)
91. ISO 19147:2015: Geographic information – transfer nodes (2015)
92. Bibliographisches Institut (ed.): Meyers großes Taschenlexikon, 9th edn. Meyers, Mannheim (2003). in German
93. ISO 19144-1:2009: Geographic information – classification systems – part 1: classification system structure (2009)
94. ISO 19152:2012: Geographic information – Land Administration Domain Model (LADM 1) (2012)
95. ISO 19160-1:2015: Geographic information – addressing – part 1: conceptual model (2015)
96. The Vienna Agreement – FAQs. https://share.ansi.org/Shared%20Documents/News%20and%20Publications/Links%20Within%20Stories/New%20ISO%20CEN%20VA%20FAQs.pdf, Accessed 16 Dec 2018
97. OGC: Bylaws of Open Geospatial Consortium. Open Geospatial Consortium, Wayland (2019)
98. OGC: The OGC Innovation Program Policies and Procedures. Open Geospatial Consortium, Wayland (2020)
99. OGC: Intellectual Property Rights Policy of the Open Geospatial Consortium, Inc. Open Geospatial Consortium, Wayland (2008). dated 12 April 2009
100. OGC: Topic 1: Features and geometry – Part 1: Feature Models. Open Geospatial Consortium, Wayland (2020)
101. OGC: Topic 2: Spatial Referencing by Coordinates. Open Geospatial Consortium, Wayland (2021)
102. OGC: Topic 3: Locational Geometry Structures. Open Geospatial Consortium, Wayland (1999)
103. OGC: Topic 4: Stored Functions and Interpolation. Open Geospatial Consortium, Wayland (1999)
104. OGC: Topic 5: Features. Open Geospatial Consortium, Wayland (2009)
105. OGC: Topic 6: Schema for Coverage Geometry and Functions. Open Geospatial Consortium, Wayland (2006)
106. OGC: Topic 7: Earth Imagery. Open Geospatial Consortium, Wayland (2004)
107. OGC: Topic 8: Relationships Between Features. Open Geospatial Consortium, Wayland (1999)
108. OGC: Topic 10: Feature Collections. Open Geospatial Consortium, Wayland (1999)
109. OGC: Topic 12: The OpenGIS Service Architecture. Open Geospatial Consortium, Wayland (2001)
110. OGC: Topic 13: Catalog Services. Open Geospatial Consortium, Wayland (1999)
111. OGC: Topic 14: Semantics and Information Communities. Open Geospatial Consortium, Wayland (1999)
112. OGC: Topic 15: Image Exploitation Services. Open Geospatial Consortium, Wayland (2000)
113. OGC: Topic 16: Image Coordinate Transformation Services. Open Geospatial Consortium, Wayland (2000)
114. OGC: Topic 17: Location Based Services. Open Geospatial Consortium, Wayland (2000)
115. OGC: Topic 18: Geospatial Digital Rights Management Reference Model. Open Geospatial Consortium, Wayland (2007)
116. ISO 19148:2012: Geographic Information – Linear referencing (2012)
117. OGC: Topic 20: Observations and Measurements. Open Geospatial Consortium, Wayland (2010)
118. OGC: Topic 21: Discrete Global Grid Systems – Part 1 Core Reference System and Operations and Equal Area Earth Reference System. Open Geospatial Consortium, Wayland (2021)
119. OGC: OGC Abstract Specification Topic 22 – Core Tiling Conceptual and Logical Models for 2D Euclidean Space, Open Geospatial Consortium, Wayland (2020)
120. OGC: OGC Implementation Standards. Open Geospatial Consortium, Wayland (2011)
121. OGC: OpenGIS Implementation Specification for Geographic Information – Simple Feature Access – Part 1: Common Architecture. Open Geospatial Consortium, Wayland (2011)

122. OGC: OpenGIS Implementation Specification for Geographic Information Simple Feature Access – Part 2: SQL Option. Open Geospatial Consortium, Wayland (2010)
123. OGC: OpenGIS Simple Features Implementation Specification for CORBA. Open Geospatial Consortium, Wayland (1999)
124. OGC: OpenGIS Simple Features Implementation Specification for OLE/COM. Open Geospatial Consortium, Wayland (1999)
125. OGC: OpenGIS Web Map Service (WMS) Implementation Specification. Open Geospatial Consortium, Wayland (2006)
126. OGC: OpenGIS Styled Layer Descriptor (SLD) Profile of the Web Map Service Implementation Specification. Open Geospatial Consortium, Wayland (2007)
127. OGC: OpenGIS Symbology Encoding Implementation Specification. Open Geospatial Consortium, Wayland (2006)
128. OGC: OpenGIS Web Map Context Implementation Specification. Open Geospatial Consortium, Wayland (2005)
129. OGC: OGC OWS Context Conceptual Model. Open Geospatial Consortium, Wayland (2014)
130. OGC: OGC OWS Context Atom Encoding Standard. Open Geospatial Consortium, Wayland (2014)
131. OGC: OGC OWS Context GeoJSON Encoding Standard. Open Geospatial Consortium, Wayland (2017)
132. Internet Engineering Task Force: The GeoJSON format (2016). https://tools.ietf.org/html/rfc7946
133. OGC: OpenGIS Web Map Tile Service Implementation Standard. Open Geospatial Consortium, Wayland (2010)
134. OGC: OGC Web Feature Service (WFS) 2.0 Interface Standard. Open Geospatial Consortium, Wayland (2010)
135. OGC: OGC Filter Encoding 2.0 Encoding Standard with Corrigendum. Open Geospatial Consortium, Wayland (2014)
136. OGC:. http://www.opengeospatial.org/standards/wcs
137. OGC: OGC Web Coverage Service (WCS) 2.1 Interface Standard. Open Geospatial Consortium, Wayland (2018)
138. OGC: OGC Web Coverage Service (WCS) Interface Standard –Transaction Extension. Open Geospatial Consortium, Wayland (2016)
139. OGC: Web Coverage Service (WCS) Interface Standard – Range Subseting Extension. Open Geospatial Consortium, Wayland (2014)
140. OGC: OGC Web Coverage Service (WCS) Interface Standard – Scaling Extension. Open Geospatial Consortium, Wayland (2014)
141. OGC: OGC Web Coverage Service (WCS) Interface Standard – CRS Extension. Open Geospatial Consortium, Wayland (2014)
142. OGC: Web Coverage Service (WCS) Interface Standard – Interposation Extension. Open Geospatial Consortium, Wayland (2014)
143. OGC: OGC Web Coverage Service (WCS) Interface Standard – Processing Extension. Open Geospatial Consortium, Wayland (2014)
144. OGC: OGC Web Coverage Service 2.0 Interface Standard – KVP Protocol Binding Extension. Open Geospatial Consortium, Wayland (2013)
145. OGC: OGC Web Coverage Service (WCS) Interface Standard – XML/POST Protocol Binding Extension. Open Geospatial Consortium, Wayland (2010)
146. OGC: OGC Web Coverage Service 2.0 Interface Standard – XML/SOAP Protocol Binding Extension. Open Geospatial Consortium, Wayland (2010)
147. OGC: OGC Coverage Implementation Schema with Corrigendum. Open Geospatial Consortium, Wayland (2019)
148. OGC: OGC Coverage Implementation Schema – ReferencableGridCoverage Extension with Corrigendum. Open Geospatial Consortium, Wayland (2019)
149. OGC: OGC GML Application Schema – Coverages – GeoTIFF Coverae Encoding Profile. Open Geospatial Consortium, Wayland (2014)
150. OGC: OGC GML Application Schema – Coverages JPEG2000 Coverage Encoding Extension. Open Geospatial Consortium, Wayland (2015)
151. OGC: CF-netCDF 3.0 encoding using GML Coverage Application Schema. Open Geospatial Consortium, Wayland (2015)
152. OGC: Web Coverage Processing Service (WCPS) Language Interface Standard. Open Geospatial Consortium, Wayland (2021)
153. OGC: OGC Web Service Common Implementation Specification. Open Geospatial Consortium, Wayland (2010)
154. OGC: OGC WPS Interface Standard Corrigendum 2. Open Geospatial Consortium, Wayland (2018)
155. OGC: OGC Publish/Subscribe Interface Standard 1.0 – Core. Open Geospatial Consortium, Wayland (2016)
156. OGC: OGC Publish/Subscribe Interface Standard 1.0 – SOAP Protocol Binding Extension. Open Geospatial Consortium, Wayland (2016)
157. OGC: OGC Geography Markup Language (GML) Encoding Standard. Open Geospatial Consortium, Wayland (2016)
158. OGC: OGC City Geographic Markup Language (CityGML) Part 1: Conceptual Model Standard. Open Geospatial Consortium, Wayland (2021)
159. OGC: OGC Geoscience Markup Language 4.1 (GeoSciML). Open Geospatial Consortium, Wayland (2017)
160. OGC: OGC IndoorGML 1.1. Open Geospatial Consortium, Wayland (2020)
161. OGC: OGC Land and Infrastructure Conceptual Model Standard (LandInfra). Open Geospatial Consortium, Wayland (2016)
162. OGC: OGC InfraGML 1.0: Part 0 – LandInfra Core – Encoding Standard. Open Geospatial Consortium, Wayland (2017)
163. OGC: OGC InfraGML 1.0 – Part 1 – LandInfra Land Features – Encoding Standard. Open Geospatial Consortium, Wayland (2017)
164. OGC: OGC InfraGML 1.0 – Part 2 – LandInfra Facilities and Projects – Encoding Standard. Open Geospatial Consortium, Wayland (2017)
165. OGC: OGC InfraGML 1.0 – Part 3 – LandInfra Alignments – Encoding Standard. Open Geospatial Consortium, Wayland (2017)
166. OGC: OGC InfraGML 1.0 – Part 4 – LandInfra LandInfra Roads – Encoding Standard. Open Geospatial Consortium, Wayland (2017)
167. OGC: OGC InfraGML 1.0 – Part 5 – LandInfra Railways – Encoding Standard. Open Geospatial Consortium, Wayland (2017)
168. OGC: OGC InfraGML 1.0 – Part 6 – LandInfra Land LandInfra Survey – Encoding Standard. Open Geospatial Consortium, Wayland (2017)
169. OGC: OGC InfraGML 1.0 – Part 7 – LandInfra LandInfra Land Division – Encoding Standard. Open Geospatial Consortium, Wayland (2017)
170. OGC: OGC GML in JPEG 2000 for Geographic Encoding Standard. Open Geospatial Consortium, Wayland (2018)
171. OGC: OpenGIS Catalogue Services Specification. Open Geospatial Consortium, Wayland (2007)
172. OGC: CSW-ebRIM Registry Service – Part 1: ebRIM profile of CSW. Open Geospatial Consortium, Wayland (2008)
173. OGC: OpenGIS Catalogue Services Specification 2.0.2 – ISO Metadata Application Profile: Corrigendum. Open Geospatial Consortium, Wayland (2017)
174. OGC: OGC Catalogue Services Standard 2.0 Extension Package for ebRIM Application Profile: Earth Observation Products. Open Geospatial Consortium, Wayland (2010)
175. OGC: OGC Catalogue Services 3.0 – General Model. Open Geospatial Consortium, Wayland (2016)
176. OGC: OGC Catalogue Services 3.0 Specification – HTTP Protocol Binding. Open Geospatial Consortium, Wayland (2016)

177. OGC: OGC Ordering Services for Earth Observation Products Interface Standard (OSEO). Open Geospatial Consortium, Wayland (2012)
178. OGC: OGC OpenSearch Geo and Time Extensions. Open Geospatial Consortium, Wayland (2014)
179. OASIS: OASIS OpenSearch – searchRetrieve: Part 4. ADP Binding for OpenSearch Version 1.0 (2013). http://docs.oasis-open.org/search-ws/searchRetrieve/v1.0/os/part4-opensearch/searchRetrieve-v1.0-os-part4-opensearch.html
180. OGC: OGC OpenSearch Extensions for Earth Observation. Open Geospatial Consortium, Wayland (2019)
181. OGC: OGC KML 2.3. Open Geospatial Consortium, Wayland (2015)
182. OGC: OpenGIS Location Services (OpenLS): Core Services. Open Geospatial Consortium, Wayland (2008)
183. OGC: OpenGIS Coordinate Transformation Service Implementation Specification. Open Geospatial Consortium, Wayland (2001)
184. OGC: OGC GeoAPI 3.0.1 Implementation Standard. Open Geospatial Consortium, Wayland (2015)
185. OGC: Geospatial eXtensible Access Control Markup Language (GeoXACML) Version 1 Corrigendum. Open Geospatial Consortium, Wayland (2011)
186. OASIS: eXtensible Access Control Markup Language (XACML) Version 2.0 (2005). www.oasis-open.org/committees/xacml/
187. OGC: OGC Network Common Data Form (NetCDF) Core Encoding Standard. Open Geospatial Consortium, Wayland (2011)
188. OGC: NetCDF Binary Encoding Extension Standard: NetCDF Classic and 64-bit Offset Format. Open Geospatial Consortium, Wayland (2011)
189. OGC: CF-netCDF3 Data Model Extension Standard. Open Geospatial Consortium, Wayland (2013)
190. OGC: OGC Network Common Data Form (netCDF) NetCDF Enhanced Data Model Extension Standard. Open Geospatial Consortium, Wayland (2012)
191. OGC: Observations and Measurements – XML Implementation. Open Geospatial Consortium, Wayland (2011)
192. OGC: Timeseries Profile of Observations and Measurements. Open Geospatial Consortium, Wayland (2016)
193. OGC: OGC Earth Observation Metadata profile of Observation & Measurements, Open Geospatial Consortium, Wayland (2016)
194. OGC: TimeseriesML 1.0 -XML Encoding of the Timeseries Profile of Observation and Measurements. Open Geospatial Consortium, Wayland (2016)
195. OGC: OGC WaterML 2.0: Part 1 – Timeseries. Open Geospatial Consortium, Wayland (2014)
196. OGC: OGC WaterML 2: Part 2 – Ratings, Gauges and Sections. Open Geospatial Consortium, Wayland (2016)
197. OGC: OGC WaterML 2: Part 3 – Surface Hydrology Features (HY_Features) – Conseptual Model. Open Geospatial Consortium, Wayland (2018)
198. OGC: OGC WaterML 2: Part 4 – GroundWaterML 2 (GWML2). Open Geospatial Consortium, Wayland (2017)
199. OGC: OGC SWE Common Data Model Encoding Standard. Open Geospatial Consortium, Wayland (2011)
200. OGC: OpenGIS SWE Service Model Implementation Standard. Open Geospatial Consortium, Wayland (2011)
201. OGC: OGC SensorML: Model and XML Encoding Standard. Open Geospatial Consortium, Wayland (2014)
202. OGC: OGC Sensor Observation Service Interface Standard. Open Geospatial Consortium, Wayland (2012)
203. OGC: OGC Sensor Planning Service Implementation Standard (SPS). Open Geospatial Consortium, Wayland (2011)
204. OGC: OGC SensorThings API Part 1: Sensing Version 1.1. Open Geospatial Consortium, Wayland (2021)
205. OGC: OpenGIS Georeferenced Table Joining Service Implementation Standard. Open Geospatial Consortium, Wayland (2010)
206. OGC: Geographic information – Well-known text representation of coordinate reference systems. Open Geospatial Consortium, Wayland (2019)
207. OGC: OGC 3D Portrayal Service 1.0. Open Geospatial Consortium, Wayland (2017)
208. OGC: OGC Augmented Reality Markup Language (ARML). Open Geospatial Consortium, Wayland (2015)
209. OGC: Primer for the OGC CDB Standard: Model and Physical Data Store Struture. Open Geospatial Consortium, Wayland (2017)
210. OGC: Volume 1: OGC CDB Core Standard: Model and Physical Data Store Structure. Open Geospatial Consortium, Wayland (2017)
211. OGC: Volume 3: OGC CDB Terms and Definitions. Open Geospatial Consortium, Wayland (2017)
212. OGC: CDB Multi-Spectral Imagery Extension. Open Geospatial Consortium, Wayland (2018)
213. OGC: OGC Geopackage Encoding Standard. Open Geospatial Consortium, Wayland (2020)
214. OGC: OGC GeoSPARQL – A Geographic Query Language for RDF Data. Open Geospatial Consortium, Wayland (2012)
215. OGC: OGC GeoRSS Encoding Standard. Open Geospatial Consortium, Wayland (2017)
216. OGC: OGC GeoSMS Standard – Core. Open Geospatial Consortium, Wayland (2012)
217. OGC: OGC Geospatial User Feedback Standard: Conceptual Model. Open Geospatial Consortium, Wayland (2016)
218. OGC: OGC Geospatial User Feedback Standard: XML Encoding Extension. Open Geospatial Consortium, Wayland (2016)
219. OGC: OGC Indexed 3d Scene Layer (I3S) and Scene Layer Package Format Specification. Open Geospatial Consortium, Wayland (2020)
220. OGC: OGC Moving Features Encoding Part 1: XML Core. Open Geospatial Consortium, Wayland (2019)
221. OGC: OGC Moving Features Access. Open Geospatial Consortium, Wayland (2017)
222. OGC: OGC Moving Features Encoding Extension: Simple Comma Separated Values (CVS). Open Geospatial Consortium, Wayland (2015)
223. OGC: OGC Open Modelling Interface (OpenMI) Interface Standard. Open Geospatial Consortium, Wayland (2014)
224. OGC: 3D Tiles. Open Geospatial Consortium, Wayland (2019)
225. OGC: CityJSON. Open Geospatial Consortium, Wayland (2021)
226. OGC: OGC EO Dataset Metadata GeoJSON(-LD) Encoding Standard EO_Geo Json, Open Geospatial Consortium, Wayland (2020)
227. OGC: OGC GeoTIFF Standard, Open Geospatial Consortium, Wayland (2019)
228. OGC: OGC Hierarchical Data Format Version 5 (HDV5®) Core Standard, Open Geospatial Consortium, Wayland (2019)
229. OGC: OGC API – Common – Part 1: Core (Draft), Open Geospatial Consortium, Wayland (2021)
230. OGC: OGC API – Common – Part 2: Geospatial Data (Draft), Open Geospatial Consortium, Wayland (2021)
231. OGC: OGC API – Environmantal Data Retrieval Standard, Open Geospatial Consortium, Wayland (2021)
232. OGC: OGC API – Features – Part 1: Core, Open Geospatial Consortium, Wayland (2019)
233. OGC: OGC API – Features – Part 2: Coordinate Reference Systems by Reference, Open Geospatial Consortium, Wayland (2020)
234. OGC: OGC PipelineML Conceptual and Encoding Model Standard, Open Geospatial Consortium, Wayland (2019)
235. OGC: OGC Web Services Security, Open Geospatial Consortium, Wayland (2019)

236. W3C: Semantic Sensor Network Ontology. W3C, Cambridge (2017). W3C Recommendation, 2017 10 19
237. OGC: OGC Symbology Conceptual Model: Core Part. Open Geospatial Consortium, Wayland (2020)
238. W3C: Time ontology in OWL. W3C, Cambridge (2020). W3C Recommendation, 2020 03 26
239. OGC: Two Dimensional Tile Matrix Set. Open Geospatial Consortium, Wayland (2019)
240. OGC: LAS Specification 1.4 OGC Community Standard. Open Geospatial Consortium, Wayland (2018)

Further Reading

241. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W.: Geographic Information, Systems and Science. Wiley, Hoboken (2001)
242. Volpe, J.A.: Key Practices for Implementing Geospatial Technologies for a Planning and Environment Linkages (PEL) Approach. US Department of Transportation, Washington (2008)
243. Robinson, J., Perkins, K.D.: Best practices for collecting geographic data in the field (2008). http://www.flmnh.ufl.edu/herbarium/methods/Georeferencingbestpractices.htm, Accessed 11 Nov 2018. University of Florida Herbarium (FLAS), Beta Version 2
244. Federal Transit Administration: Best Practices for Using Geographic Data in Transit: A Location Referencing Guidebook Defining Geographic Locations of Bus Stops, Routes and Other Map Data for ITS, GIS and Operational Efficiencies. Rep. Number FTA-NJ-26-7044-2003.1. US Department of Transportation, Washington (2005)
245. ISO: Report of the JTC1 Ad Hoc Meeting of the Technical Direction on Cultural and Linguistic Adaptability and User Interfaces. ISO/IEC JTC1 document N5629. (1998)
246. ISO/IEC TR 11017:1998: Information technology – Framework for internationalization (1998)
247. ISO 639-2:1998: Codes for the representation of names of languages – Part 2: Alpha-3 code (1998)
248. ISO/IEC 11179-3:2013: Information technology – Metadata registries (MDR) – Part 3: Registry metamodel and basic attributes (2013)
249. Ghiladi, V.: Do standards users get what they want from ISO standards? Iso Bull. **33**(11), 2 (2002)
250. Anold, D.B., Duce, D.A.: ISO Standards for Computer Graphics, The First Generation. Butterworths, London (1990)
251. ISO/TS 19139:2007: Geographic information – Metadata – XML schema implementation (2007)

Wolfgang Kresse Wolfgang Kresse holds a Diploma degree in Geodesy and a Doctor's degree in Digital Cartography, both from the University of Bonn, Germany. He was Professor of GIS and Photogrammetry at the University of Applied Sciences in Neubrandenburg, Germany, and Professor of remote sensing at the Politechnika Koszalińska in Poland. As a Visiting Professor he was affiliated with the Canada Centre for Remote Sensing in Ottawa (2002/2003) and the University of Maine, USA (2008/2009). He has also worked as a software developer and customer trainer at the Carl Zeiss company with a focus on graphics, photogrammetry, and data exchange.

David Danko David Danko is Senior Consultant with Esri, where he provides professional consulting services in GIS standards, interoperability, and geographic metadata. Previously, he worked for the US Defense Mapping Agency (DMA), where he was responsible for developing and managing agency policies and programs to ensure standardization and interoperability. David Danko has over 40 years experience in the fields of cartography, photogrammetry, remote sensing, and geographic information.

Kian Fadaie Dr Fadaie was National Director of Hydrography at the Canadian Hydrographic Service and a member of the IMO e-Navigation Correspondence Group. She was responsible for standards and interoperability of hydrographic data. Dr Fadaie chaired the Canadian Advisory Committee to ISO/TC 211 for 8 years and was the proponent of the Imagery, Gridded and Coverage Data Standards in TC 211. She has authored many scientific papers, books, publications, citations and guides.

Geospatial Web Services

16

Joan Masó

Contents

Abstract

Geospatial web services are based on the communication networks that connect computers on the Internet. The communication network protocols are structured in layers. Top layers hide the incremental complexity of the heterogeneous hardware. Geospatial web services can be implemented using Hypertext Transfer Protocol (HTTP) to distribute geospatial resources (resource-oriented architecture style) or can be implemented as a new layer of protocol on top of HTTP (service-oriented architecture style). Geospatial web services separate the application into a client side that requests information and a server side that responds to that request, using a common communication protocol (that specifies the communication language and request and respond formats). These two sides might be developed using different programming languages and might be implemented by several vendors; however they are able to interoperate because they use a set of well-defined internationally agreed open standards.

Geospatial web services can be classified based on their function: visualization services, data exchange services, processing services and discovery services. Even if different standards deal with different categories, they respect a common set of characteristics making possible the creation of integrated clients where all service categories are used to fulfill user needs.

This chapter provides an introduction and background for geospatial web services that is later used to describe the main operations of the different categories of geospatial services in a very practical way, using examples that illustrate the communication with real services available at the time of writing this chapter.

An emerging new type of geospatial standards, based on a resource-oriented architecture style and the definition of open application program interfaces, is presented. They might provide the basis of a new generation of geospatial web services that will be more in line with mainstream information technology.

Keywords

web services · communication · networks · resources · geospatial

J. Masó (✉)
CREAF Grumets research group, CREAF. Edifici C. Universitat Autònoma de Barcelona
Bellaterra, Catalonia, Spain
e-mail: joan.maso@uab.cat

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_16

16.1 Background

Most of us use a mobile phone on a daily basis for tasks other than making phone calls. We know that most of the apps that we have installed in our phone require internet connectivity (so called "mobile data" or WiFi). These applications store a minimum set of data in our phone but also store other data in the cloud (in the network, somewhere). We have become accustomed to applications requiring external and distributed computers to work. Possibly, the oldest popular example of this kind of application is the email client. We know that each time we write and send an email, other computers will help in the task of routing it to the final email box and, at the same time, there is a remote computer that will later give us access to the incoming responses. The most popular internet application is still the web browser. In web browsers, a white page is filled with content coming from the outside (from remote web servers). When the web began, we got used to waiting a bit for the page to react to our actions and repopulate the white screen with content, and that was because we knew it was a "long-distance" communication. We also got used to the fragility of the remote computing that sometimes responded with mysterious numbers like 404 or 500, indicating some kind of failure. Web interfaces matured, and the web interaction was more fluid and the transitions between actions were smoother until a new kind of web page emerged: the web applications. They also run in web browsers but look and feel similar to desktop programs, except that they rarely save content onto our machine but into the cloud.

Focusing on the geospatial domain, Google maps popularized the use of geospatial information by removing the difficulties in dealing with large geospatial files and simplifying location-based queries to remote geospatial databases. This transformation from monolithic applications to distributed computing was made possible by separating the different parts of the applications and distributing them in separated (far apart) computers in a computer architecture called the web services architecture [1]. Web services come in two flavors: closed and open. This chapter focuses on the open web services, where the different elements of the distributed architecture are fully known, and some modules are provided many times by different vendors in an interoperable way. Many of the principles explained here can be applied to closed systems, but the fact that they are not based on open standards makes their complete description impossible.

16.2 Elements of Distributed Computing

This section provides the necessary background about network communications and web protocols. The concepts introduced in this section give us the necessary basis to describe the open geospatial web services in following sections.

16.2.1 Client–Server Architecture

A typical application usually includes three essential elements: presentation, logic, and data. The graphical user interface is the responsible element for presentation, the processing code is the responsible element for logic, and the database or the file management system is the responsible element for the data. The three elements are interconnected and, when needed, one element makes a request to another one, which fulfills the request and responds with the necessary information to the first.

The element making the request is called *the client*, and the element fulfilling the request is called *the server*. The user typically interacts with the graphical user interface (the client), and the client software is the responsible for communicating with the logic and the data (the server). Which physical machine contains the three elements depends on the architecture of the application.

- In a typical stand-alone GIS application, the three elements reside in the same computer, and you need to be a programmer to actually realize that this separation exists.
- In a typical GIS application accessing the data over a local area network, the presentation and the logic are on one machine, but the data is on another.
- In a web application, the presentation is on one machine, while the logic and the data are on another.

The last alternative is the one selected by the web services, and this will be the architecture that is mainly considered in this chapter.

We have seen that the physical role of the three elements depends on the architecture, but the logical roles can depend on the architecture as well, so the logical role of client and server is not fixed. For example, the logic element is the server from the presentation point of view, but it is the client from the data access point of view that acts as a server.

Later we are going to see that the web client–server architecture is a special case of the general client–server architecture. In the web client–server architecture, the client and server communicate using a web protocol (Fig. 16.1).

16.2.2 Messaging

Client and server are two pieces of software that need to communicate in a common agreed upon language and using a common set of rules (protocol) in such a way that they can understand each other. The role of the protocol is to connect the client and the server in a way that is transparent to users and that hides the complexity of the system. There are many connection models, such as the remote procedure call

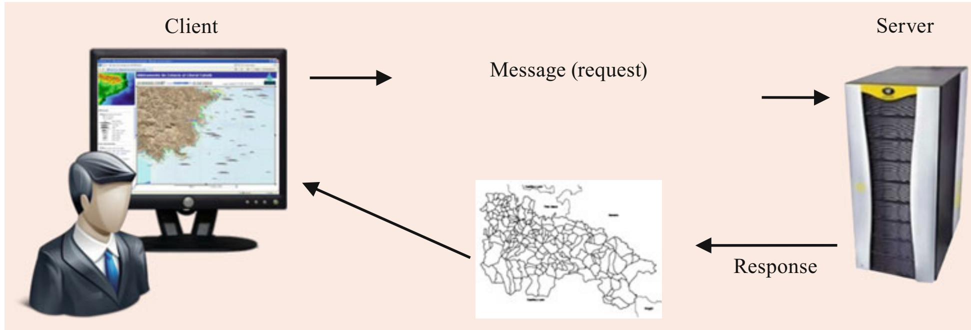


Fig. 16.1 Client–server architecture and the protocol used

(RPC) and publish/subscribe (PubSub). In an RPC, a client sends a request to a remote server and waits for the server to respond. The server executes the procedure, generates the answer, and returns the result back to the client who receives it and closes the connection. HTTP uses RPC. RPC does PULL requests because the client initiates the movement (as if the client were pulling a door to enter the communication). In PubSub (which does a PUSH request), the client subscribes to some events based on the interests of the user and, later, the server will initiate a communication to the client when there is a need for the client to be notified, depending on the client subscription preferences. One typical application in the GIS world is a client that subscribes to changes in the features in some area of interest. When changes in the scope of the subscription are made, the server notifies the client about those changes. PubSub can be implemented on top of TCP/IP. HTTP cannot do PubSub directly, but it can emulate it by forcing the client to regularly request the server if changes have occurred in its subscription, using RPC.

16.2.3 Levels of Protocol

Web services are developed over a stack of protocols organized in different levels, one on top of the other. To understand the need for a stack of protocols, an analogy can be used as follows. The bank industry had set up a communication protocol with their clients based on sending them some standardized letters where different expenditures and incomes are detailed. Clients regularly receive these messages, review them, and react accordingly. This protocol is part of a bank mail notifications protocol where commercial advertisements and other communications are performed by letters all having the bank brand logo very visible on the envelope. This protocol level uses yet another lower-level protocol of the postal correspondence. In the postal protocol, letters need to be in an envelope, with a destination postal address and a stamp. Envelopes are deposited in a post office, and they will be classified and travel a sequence of post offices until they are close enough to be delivered to the final address. This communication protocol relies on a heterogeneous very lower-level

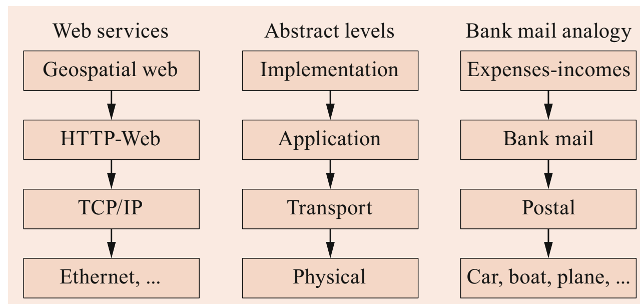


Fig. 16.2 Levels of protocol

protocol: the communications network formed by cars, boats, and planes that follow a set of rules through their own media to move around people, goods and, of course, mail (including the bank mail). The bank industry is using the whole stack of protocols without even realizing the complexity of the system. However, dependencies exist, and if something goes wrong in a low-level protocol (e.g., a truck strike), the whole stack cannot function properly.

The web protocol is a high level protocol and, below that, other layers of protocol are defined one on top of the other (Fig. 16.2). At the very bottom of the stack, the physical communication layer use a mixture of protocols orchestrating communications among the physical devices and cables of the network, such as mobile phones, Ethernet cables, fiber optic cables, wireless connections, etc., all making internet communication possible. On top of the physical layer the transport protocol uses transmission control protocol/internet protocol (TCP/IP) as a network protocol. It is based on fragments of messages called packets that are transmitted through a sequence of computers that route each packet from the source to the target computer. Each computer has an IP logical address that univocally identifies it, and the IP address is used to route the packet from one computer to the next. On top of TCP/IP, the actual web protocol is defined as hypertext transfer protocol (HTTP). HTTP is an application protocol for hypermedia information systems. Hypertext is structured text that uses logical links (hyperlinks) between documents containing text. HTTP is the protocol designed to exchange or transfer hypertext.

Upper levels of protocol depend on the lower levels; sometimes the separation between levels of protocol is not strict. We shall see *implementation protocols* that use (or misuse) some of the HTTP *application protocol* characteristics.

16.2.4 HTTP

HTTP is a request–response application protocol that is also based on the client–server architecture. The most used client application is the web browser, and a web server application running on a computer hosting a website is the most common server. The client submits an HTTP request message to the server and waits. The server returns a response message to the client, generally containing a single document, the most common being a hypertext markup language (HTML) file. The response contains a *completion* status, which is information about the request, in a section of the message that is called the *headers* and may also contain the requested content in its message *body* (e.g., the HTML file).

In HTTP, resources are identified and located on the network by uniform resource locators (URLs), using the uniform resource identifiers (URI's). URIs are the basis for requesting documents and are also the basis of the hyperlinks in HTML, acting as a way to link a fragment of an HTML page to another HTML file. A URL can represent prestored data or data that is generated dynamically; depending on the implementation on the server side. Commonly, resources correspond to preexisting files but they can be the dynamic output of an executable application residing on the server side.

HTTP requests use different methods (also called *verbs*). All methods use a URI, but depending on the method, the client can require a desired action to be performed on the identified resource on the server side. The current specification of HTTP 1.1 defines several methods that have a clear semantics on the action to be done by the server on the URL requested:

- The GET method requests (retrieves) a representation of the specified resource. Requests using GET should only retrieve data and do not contain any payload. The expected response is a document representing the resource as a body of the response.
- The HEAD method asks for the same resource than a GET request, but the response will not contain the resource but only metadata about it in the headers of the response.
- The POST method sends a resource in the body of the message that is requested to be created and stored in the server. The URL of the POST method is NOT the URL of the resource but a new subordinate of the web resource identified by the URI. The resource is expected to be saved and the URL of the new resource returned in the headers of the response.
- The PUT method sends a resource in the body of the message that is requested to be stored under the supplied URI. Generally, PUT is used for updating existing resources, but if the URI does not point to an existing resource, then the server can create the resource with that URI.
- The DELETE method requests that the resource pointed by the URL be deleted.
- The TRACE method echoes the received request so that a client can see what (if any) changes or additions have been made by intermediate servers.
- The OPTIONS method requests information about the HTTP methods that the server supports for a specified resource URL. It can be used to check if a *verb* is supported before actually requesting it.
- The PATCH method applies partial modifications to a resource.

All general-purpose HTTP servers are required to implement at least GET and HEAD methods, and all other methods are considered optional by the specification.

HTTP is a stateless protocol. In a stateless communication protocol, no information is retained by either sender or receiver. In practice, this means that requests are independent of each other, and the client should not assume that any acknowledge of any previous communication was retained by the server. The stateless design simplifies server design because there is no need to save and store any details of a request or a response while dealing with conversations between client and server. In practice, an application needs to maintain the state of the dialog with the user, so in a stateless communication, details about the state are sent to the server in every request; the server needs to interpret and, many times, return back status information to the client (before closing the individual request–response circle and losing the information) to allow for the state of the application to be maintained.

The HTTP protocol was designed to transport information in plain text mode. Due to the nature of TCP/IP, the computers in the local area network of the client and the server and all intermediate computers that route the communication have access to plain text and can easily read it or even store it. Confidential information transmitted in this mode can fall into the hands of unwanted people, including passwords, credit card numbers, etc. To prevent this a secured version of HTTP was introduced: HTTPS. In HTTPS, the URL and the message are encrypted and cannot be read by any computer but the receiver.

16.2.5 The Web Server

A web server is a software (and a hardware that executes it) that implements the server-side HTTP application protocol by processing incoming network request and response documents. The primary function of a web server is to store, process, and deliver web pages to clients. Most common web pages are HTML documents and all related content, which may include images, style sheets, scripts (e.g., JavaScript), etc. Common examples of web server software are internet information services, Apache server, Nginx, etc. [2].

A web client, commonly a web browser, initiates a communication by making a request for a specific resource URL, and the server responds with the requested document or an error message (e.g., if the URL resource does not correspond to any content). The resource might be a real file on the server storage system, but it could also be generated on the fly based on the content of the server database. This is particularly useful for serving geospatial information that is usually stored in databases that can be dynamically converted into documents (e.g., a GML or a GeoJSON document) on the fly.

The primary function is to serve content by using the GET operation, but complete implementation of HTTP also allows the server to receive content from clients (PUT and POST operations). This feature is commonly used for submitting web forms (e.g., when users answer surveys) or to upload files (e.g., when users store or exchange files on the cloud) and could be utilized to accept and store user edits on a map.

Web servers also support that a URL points to an executable. This executable will be conformal to the CGI, ISAPI, or a similar alternative. When the URL to an executable is received, it is executed and it will use the information in the request to perform a task (e.g., generate a geospatial resource). In the geospatial domain, the executable could render a map (from geospatial information in a database) in an image and return it as the response. Modern web servers store documents that contain a server-side code (such as active server pages (ASP), PHP, etc.) that will be dynamically interpreted by the server when the URL of the server script is requested. Executables and server-side scripts are means to augment the behavior of the web server by adding separated elements, while the actual web server main software remains unchanged. Later, we will see that geospatial services will use this characteristic to implement geospatial services on top of general-use web services.

16.2.6 The Web Client

A web client is a piece of software that implements the client-side HTTP application protocol by generating network requests and handling and displaying response documents. The primary function of a web client is to visualize web pages for users. The most common web client is the web browser, which is a generic application for accessing and visualizing many kinds of information available on the world wide web (WWW) on the user's device. It also has to react to human actions; the most common one being a click on a hyperlink. The web browser is able to visualize HTML pages, images, videos, etc. Modern web browsers are able to handle several kinds of information, apply styles to the HTML pages using style sheets, and execute client-side scripts (e.g., JavaScript). Common examples of web client software are Internet Explorer, Microsoft Edge, Google Chrome, Mozilla Firefox, Safari, Opera, etc.

Images embedded in HTML files were the first map applications shown in web browsers. Modern web browsers are complicated applications supporting several formats, but they can accept the inclusion of external supplementary code (e.g., with ActiveX). This method is used, for example, to add JPEG2000 imagery viewers.

In recent years, web browsers have evolved fast, but their performance cannot be compared to the desktop solutions. In addition, they have security restrictions that severely limit access to local resources (in particular to the local storage drive). To overcome those limitations, application programmers can implement or incorporate web clients in desktop solutions specifically tailored for the GIS data requirements that will be capable of combining local files on top of remote data.

16.2.7 The URL Structure

The URL is a fundamental part of the web request, and all HTTP methods use it. In the HTTP protocol, the URL has several functionalities embedded in a single query string: The URL is the way that the client has to identify the server. In addition, a URL points to a resource in the web server space; this resource can be a document or a dynamic server-side script or application. In the case that the resource is a dynamic server-side script, the URL can include a query string to the server. A URL can also contain a final fragment that the client will process after the server responds. The full URL syntax structure is a sequence of six components disposed like `<protocol>://<host>:<port>/<resource>?<query>#<fragment>`.

For example: `http://hostname.com:80/pdfs/mypdf.asp?info=houses#page=3;`

- *protocol* (in the example `http`) is the protocol name. In HTTP, it can be http or https.
- *host* (in the example `hostname.com`) is the name (or the IP address) of the server that contains the resources we are interested in. Commonly, it is composed of works separated by dots (e.g., www.maps.com).
- *port* (in the example `80`) is the TCP/IP port number. If it is missing and http protocol is indicated, 80 is assumed. If https protocol is indicated, 443 is assumed. Some map services uses other ports by default to avoid conflict with other preexisting services in the server. The Windows

distribution of GeoServer setup offers the port 8080 by default. If the default is accepted, all communications to the service need to contain `:8080` in the URL to work.

- *resource* (in the example `pdfs/mypdf.asp`) is the path and name of the resource (e.g., a HTML file to retrieve or a server-side application or a server-side script to execute).
- *query* (in the example `info=houses`) is the query information that is delivered to a server-side application or script. This query can carry parameters to inform the application or the script what to do. Key and value pair (KVP) is a particular form of query that represents a sequence of blocks consisting of a key name, an =, and a value. Each block is separated by an `&`. KVP is particularly useful if the resource is a map server that will deliver dynamic maps; this query can transport the extent, the coordinate reference system, the size of the map, the types of features it will contain, the style to portray them, etc.
- *fragment* (in the example `page=3`) is information that will be returned to the client in the response. The client side can use it to do an action. The most common use of "fragment" is to contain a name of an anchor inside a HTML page. Once the browser receives the HTML page, it will scroll down to ensure that the content after the anchor is visible. For example, Adobe PDF uses it to indicate a page number where the book should be presented when it is received and opened in the screen.

16.2.8 The HTTP Message Structure

In the interaction between client and server, a message is sent. Once the server responds, another message is sent back. These messages are plain text. The general structure of the two messages is the same. Basically, they have two parts: the header and the body. They begin with a header, which contains a sequence of text lines composed by a name, a ":" and a value. A blank line will mark the end of the header, and the body of the message will follow. The headers are generally not shown to the user and contain metadata about the communication or the body. They might contain the media type of the body, the length of the body, the calling URL, cookies, security information, language negotiation information, etc. In a request message, headers will also contain the method (e.g., GET, POST) and, in the response, headers will contain the HTTP status number. The HTTP status number is an important part of the response, which will indicate if the request has been successfully executed (e.g., by returning 200) or if something went wrong (e.g., 404 means "file not found" and 500 means "error in a server-side application"). The body of the message contains the actual payload. In a GET request, the message has an empty body, and the response should contain the document requested in the URL or the result of the execution of an application pointed by the URL. In a POST or a PUT request, the body will contain information for the server to save. Even if the HTTP message is plain text, there are ways to transmit a binary payload by reencoding it in text using one of the common methods available (initially designed for email attachments, e.g., base64). For example, a map server need will use base64 to encode the body of the response containing a map renderization in a JPEG file format that the client has requested with the intention of presenting it on the screen.

16.2.9 SOA and ROA Services on Top of a Web Server

We have explained that executables and server-side scripts are means to augment the behavior of the web server. Most of the web services implemented on top of HTTP use this approach.

Many services follow the *service-oriented architecture* (SOA) style. In this approach, services implement a new level of protocol on top of HTTP, called the *implementation level*. These services expose a collection of operations that are executed by the service and define how a client can send a message to the server and what the server will respond. A typical minimum set of operations will be an operation to request information about what the server can do and another operation to retrieve data from the service. Other services might include operations to upload data from the user to the server (known as transactional services). Since a SOA application is implemented as a new protocol layer, it can ignore the strict semantics of the HTTP methods and will define its own methods and operations in the new level.

Another way of building web services is called *resource-oriented architecture* (ROA) style. These services use the HTTP protocol methods (POST, GET, PUT, and DELETE) to formulate basic operations (create, retrieve, update, delete, respectively) on resources in the server and focus on the definition of the resources beyond the HTML format. How these resources are going to be communicated between clients and servers is predefined by the HTTP protocol in the application level. ROA services are modernly referred as Web APIs.

SOA services enjoyed popularity in the past 15 years due to their capability to function on top of HTTP and their completeness. An SOA service defines every necessary detail of the implementation level protocol. Despite SOA's popularity, web APIs and ROA services have been gaining adepts in the recent years. The main advantage of ROA services is the benefits from the continuous improvements that W3C is implementing in the HTTP and HTTPS protocols and the fast implementation in the main web browsers vendors. An ROA service will immediately incorporate the new features in HTTP or HTTPS in terms of security and reliability, while a SOA service is forced to consider these aspects in the protocol implementation level by itself. However, in the opinion of the author of this chapter, a pure ROA style might be too simplistic for complex geospatial operations such as the conflation of two datasets. Discussions on the need to associate all geospatial information to the resource concept have delayed the implementation of geospatial ROA standards considerably.

16.2.10 RESTful Web Services

With the huge success of HTTP, people have become interested in mimicking how the web works when implementing other web services. In his PhD dissertation, Roy Fielding, one of the main authors of the HTTP specification, defined the representational state transfer (REST) architectural style as the basis on how the web works [3]. Services that are implemented following the web mechanisms are called RESTful services. There is still some discussion on what the exact level of similarity to the web is that should be considered to be a truly RESTful service, but the generic consensus is that it starts by considering a resource-oriented architecture over HTTP (as described in the previous section), but it also needs to use hypermedia documents. In RESTful services, hypermedia is the engine of the application state (HATEOAS), meaning that the client application is able to understand hypermedia formats (e.g., HTML files) and should dynamically navigate to the appropriate resources by traversing the hypermedia links (aka using links in a resource to generate new requests to other resources). In practice, this means that resources need to be explicitly linked together in a network.

One of the characteristics of geospatial information that makes it unique is georeferencing. By georeferencing data it becomes spatially (and temporally) related, and implicit geospatial relations emerge (colocation, overlap, neighborhood, etc.). We can assume that both clients and servers are geospatially aware and can make use of these relations. For example, given a cell (pixel) of a coverage, there is always a cell immediately to the north, and others immediately to the south, east, and west. The same is true for 2-D tiles. To make these relations explicitly links are possible but extremely verbose and sometimes inefficient. A geospatial client (and server) can have these relations hardcoded and can use them, removing the need to communicate them. In some other cases, explicit relations could be useful; it could be worth expressing topological relations among polygons (e.g., contiguity) as links, which could otherwise be time-consuming to recalculate. Generally speaking, HATEOAS is not appropriate or necessary for geospatial information, and, in the opinion of the author of this chapter; RESTful services cannot be fully applied to all geospatial domains.

16.2.11 First Attempts to Visualize Maps in HTML

HTML was initially defined to support hypermedia. To make the presentation more visually appealing, GIF, JPEG, or PNG images were included. Soon, it became apparent that a server could be configured to automatically create maps by generating GIFs or JPEGs. In June 1993, the Xerox Corp. generated what is considered as the first interactive map viewer. Based on the selection of the user on a web page, the client software requested maps to a service: an application that was able to retrieve geospatial data from a database and render it at screen resolution and convert it into a GIF file that was returned to the client to be included as an illustration on a HTML page. Many other map browser solutions started to appear, such as the open source MapServer, and many proprietary web pages started to operate, such as MapQuest, all using the same principle. However, none of them could communicate with any of the other ones.

In 1997, the Open GIS Consortium (later renamed as Open Geospatial Consortium; OGC) proposed the OGC WWW mapping framework with the mission to investigate how to make similar mapping approaches implemented by different vendors interoperable. In 2000, the OGC proposed the version 1.0.0 of the web map service (WMS). WMS has been a huge success, and today there are thousands of standard web map services in the world produced by governments, academia, and the commercial sector, and hundreds of web map browsers [4].

The OGC standardized other services that exchange other kinds of geospatial information and adopted XML and XML schema as the *lingua franca* to encode many media types, for example service metadata, feature-based data (GML), observations-based data (O&M), etc. None of the other OGC services has become so popular or has had the same impact as WMS, which may be due to the fundamental limitations in the early versions of HTML.

16.2.12 JavaScript

JavaScript started as a Netscape's cross platform, object-based scripting language for web applications. Currently, most popular web browsers support it and some provide mature debugging tools. Web browsers interpret JavaScript statements embedded in an HTML page and execute them on the fly. The browser reads the HTML page from top to bottom, displaying the results of the HTML and executing JavaScript statements as they are encountered. JavaScript statements embedded in an HTML page can also respond to user events, such as mouse clicks, form inputs, and page navigation. For example, a JavaScript function can verify that a user has entered valid information into a form requesting a latitude and longitude. Without any network transmission, the embedded JavaScript on the HTML page can check the data entered and display a dialog box if the user enters invalid data. Modern JavaScript libraries can contain thousands of lines of code and can be used together in a single application. JavaScript can request more data to the service while the HTML page continues responding to the user in a strategy called Asynchronous JavaScript and XML (AJAX). For example, a JavaScript application can request from a WFS service more feature data and interpret the response in the form of a GML file, while the user can continue interacting with the map browser.

Web-based map clients are coded in JavaScript such as OpenLayers and Leaflet and provide JavaScript APIs with a vast list of functions for developers for presenting and manipulating geospatial data on HTML pages easily.

Recently, a small subset of the JavaScript language called JavaScript object notation (JSON) has emerged as a simpler and easier to use alternative to XML. JSON is not a programming language but a way to encode objects in JavaScript that has been made independent from it and can be used elsewhere. JSON does not include the concept of classes and data types, but objects can be validated under a certain rules defined in a JSON Schema that acts similarly to the way that XML schema does for XML. Another recent advantage of JSON objects is that they can use JSON-LD to be translated to RDF automatically by applying JSON context documents [5].

16.2.13 Improvements in HTML5

HTML5 is the last effort to make HTML a better encoding for developing applications on mobile platforms. Listing all the improvements of HTML is out of the scope of this chapter, but there are two new characteristics that are worth mentioning.

HTML5 introduces a new graphics programming interface called *Canvas*. The `<canvas>` element is a region of the HTML where JavaScript programmers have full control of the graphical content. It is possible to extract a raster representation of the canvas and change pixel colors one by one. It also incorporates functions to draw icons, lines, and polygons. We foresee that canvas will enable a completely new generation of geospatial applications, able to present vector data directly and to dynamically change visualization styles for both raster and vector data [6].

16.3 Geospatial Web Services

Geospatial data is constantly being created by cartographic agencies, automatic in-situ sensors, remote sensing satellites, etc. In many countries, it is in the aim of a transparent administration to share and exchange data with citizens, who can then formulate informed opinions. In addition, governments have recognized that by making data public, private companies can discover new usages that go beyond the ones initially foreseen by their producers and generate new business models and profits by selling new products and services. In this chapter, we shall review current geospatial web services that facilitate the dissemination of geospatial data over the web [7].

Most of these geospatial services have been defined by the OGC community, based on a consensus process that favors interoperable geoprocessing and provides services to communicate to remote geodatabases. In principle, OGC standards are defined for multiple distributed computing platforms while maintaining common geospatial semantics across the underlying technology. In practice, most OGC geospatial services use the internet infrastructure and the HTTP protocol in the SOA (and sometimes ROA) architectural style described in Sect. 16.2.8.

16.3.1 Service Types

One might expect that a single big web service would be able to deal with all geospatial aspects by incorporating all the necessary operations, in the same way that a desktop GIS in your computer is able to perform most of the necessary operations. There have been some successful attempts to do this in the commercial sector (such as the Esri ArcGIS REST API), however, this is not a common situation. Standard geospatial web services deal with a single aspect and a single data model. If we want to make use of this variety of alternative services, we need to classify them. The OGC reference model proposes a way of classifying geospatial web services by purpose and the later by the data model used [8]. This chapter follows the same approach:

- Visualization services have the purpose of presenting the data in an easy to understand way. Clients are able to request formats that can be easily placed on a computer screen by client software, such as JPEG, PNG, or KML. Servers will be able to respond with visualization formats on demand. To be able to do this, visualization services need to combine geospatial data with symbolization rules to create a presentation for the geospatial data.
- Data exchange services have the purpose of giving immediate access to data and, in some cases, permitting the upload of new geospatial data to the server. Clients should be able to request geospatial data exchange formats such as GML, GeoJSON, GeoTIFF, or NetCDF. Servers need to respond with the subset of data (as requested), and clients will be able to import the data in their GIS data storage systems. In some cases, clients should also be able

to send geospatial features to the server that should store them for further use.

- Data scheduling and notification services have the purpose of preparing the future acquisition and notifying when it becomes available in the data exchange services. The client should be able to specify the use needs in terms of extent, time, data type, etc. The server will schedule the acquisition and issue a notification when new data becomes available.
- Discovery services have the purpose of storing metadata about the data and services available on the web. Clients will formulate a request describing the type of data they need in terms of extent, projection, resolution, data quality, theme, etc. Servers should be able to analyze the query and respond with a set of metadata records describing the data or services that match the needs of the client.
- Processing services offer geospatial processing routines and hardware. Clients should be able to formulate a request that includes the processing operation, the data to be processed and the parameters to be used to tune the process execution. Servers should be able to perform the requested task and respond with the result of the operation, which will commonly be in the form of a geospatial dataset.

16.3.2 Common Architecture for Geospatial Web Services

As discussed before, the geospatial web services are separated into different types that follow standards developed by the consensus of different groups. Fortunately, there are mainly two common functionalities that all services implement: a common mechanism for formulating requests and a service self description in the form of service metadata (the GetCapabilities operation). These and other minor common characteristics are defined in the OGC Web Service Common standard [9].

In SOA style, OWS common also defines one or more common encodings to formulate requests to the service-based operations. These operations are defined in the standard and are enumerated in the service metadata document. The first encoding is the KVP on HTTP GET URLs. In the query part of the URL, a request identifier, a service name and a version number can be specified. Depending on the request name, other keys defined in the corresponding standard can be expected. Sometimes, operation parameter values are too complex to be included in a URL (e.g., when an XML file needs to be sent to the server) and XML-POST can be an alternative. In these cases, HTTP POST is used to send an XML document containing all parameters needed as instructions to execute the operation on the server. OGC services also support another encode that looks similar to the XML-POST but uses an IT mainstream approach, SOAP. In SOAP, requests are embedded in XML files called envelopes. SOAP services publish another self-discovery document called WSDL. Modern OGC web services are defined in an abstract level that can be implemented using one or more of the encoding alternatives described: KVP, XML-POST, and SOAP.

A self-describing service provides enough information for a client to discover the resources available and the capacities of the service without user intervention. In OWS Common, this is done by specifying that all services will provide a service metadata document. The OGC uses a simplified version of the ISO 19115-1 metadata for services. A service metadata document should contain the following information:

- Identification: contains metadata to identify this server in terms of standards and profiles followed, keywords describing the general content, and fees and access constraints.
- Provider: contains information about the organization operating this service.

In addition, if the service uses the SOA style, it will include

- Operations that enumerate the operations that this service supports. One of the operations is for retrieving the service metadata document (GetCapabilities).
- Resources: that enumerate resource identifiers that this service offers.

All SOA-style services implement an operation to retrieve the service metadata. This operation is always available as a KVP URL that has the query part composed mainly of three important keys: `Service` (containing the name of the service, e.g., WCS), `request` (which will contain the fixed text GetCapabilities), and `AcceptVersions` (indicating a list of version that the client understands). Some old service standards use `version` with a single version number instead of `AcceptVersions`. An example of a GetCapabilities request will look as follows: `https://acdisc.gesdisc.eosdis.nasa.gov/daac-bin/wcsL3?service=WCS&version=1.0.0&request=GetCapabilities`.

By default, the response of a GetCapabilities is in XML format. In the example (Fig. 16.3), the first subelement `wcs:Service` contains the information about services identification and provider. The second subelement `wcs:Capability` enumerates the available operations in the service showing the GetCapabilies and Describe Coverage operations (the list has been shortened for readability). The third subelement `wcs:ContentMetadata` enumerates the available coverage names with a minimum set of additional metadata for each one.

16

Fig. 16.3 Service metadata document for a WCS

```
<?xml version="1.0" encoding="UTF-8"?>
<wcs:WCS_Capabilities version="1.0.0" updateSequence="2012052200"
xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:gml="http://www.opengis.net/gml"
xmlns:wcs="http://www.opengis.net/wcs"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.opengis.net/wcs
http://schemas.opengis.net/wcs/1.0.0/wcsCapabilities.xsd">
    <wcs:Service>
        <wcs:description>This Web Coverage Service (WCS) is one of the multiple
GES DISC data service instances used to provide gridded Level 3 daily OMI data
products.</wcs:description>
        <wcs:name>Ozone Monitoring Instrument (OMI) data from NASA Goddard Earth
Sciences Data and Information Services Center (GES DISC)</wcs:name>
        <wcs:label>Ozone Monitoring Instrument (OMI) data from NASA Goddard Earth
Sciences Data and Information Services Center (GES DISC)</wcs:label>
        <wcs:keywords>
            <wcs:keyword>Aura OMI WCS NASA GSFC DISC</wcs:keyword>
        </wcs:keywords>
        <wcs:responsibleParty>
            <wcs:individualName>Long Pham</wcs:individualName>
            <wcs:organisationName>DAAC/GSFC/NASA</wcs:organisationName>
            <wcs:contactInfo>
                <wcs:phone>
                   <wcs:voice>1 301 614 5132</wcs:voice>
                </wcs:phone>
                <wcs:address>
                   <wcs:deliveryPoint>NASA/GSFC Code 610.2</wcs:deliveryPoint>
                   <wcs:city>Greenbelt</wcs:city>
                   <wcs:administrativeArea>Maryland</wcs:administrativeArea>
                   <wcs:postalCode>20706</wcs:postalCode>
                   <wcs:country>USA</wcs:country>
                   <wcs:electronicMailAddress>long.b.pham@nasa.gov
                       </wcs:electronicMailAddress>
                </wcs:address>
            </wcs:contactInfo>
        </wcs:responsibleParty>
        <wcs:fees>NONE</wcs:fees>
        <wcs:accessConstraints>NONE</wcs:accessConstraints>
    </wcs:Service>
    <wcs:Capability>
        <wcs:Request>
            <wcs:GetCapabilities>
                <wcs:DCPType>
                   <wcs:HTTP>
                       <wcs:Get>
                           <wcs:OnlineResource
xlink:href="http://acdisc.sci.gsfc.nasa.gov/daac-bin/wcsL3?"/>
                       </wcs:Get>
                   </wcs:HTTP>
                </wcs:DCPType>
            </wcs:GetCapabilities>
            <wcs:DescribeCoverage>
                <wcs:DCPType>
                   <wcs:HTTP>
                       <wcs:Get>
                           <wcs:OnlineResource
xlink:href="http://acdisc.sci.gsfc.nasa.gov/daac-bin/wcsL3?"/>
                       </wcs:Get>
                   </wcs:HTTP>
                </wcs:DCPType>
            </wcs:DescribeCoverage>
        </wcs:Request>
        <wcs:Exception>
            <wcs:Format>text/xml</wcs:Format>
        </wcs:Exception>
    </wcs:Capability>
    <wcs:ContentMetadata version="1.0.0">
        <wcs:CoverageOfferingBrief>
            <wcs:description>
```

Fig. 16.3 (Continued)


```
Spectral Absorbing Aerosol Optical Thickness for best fit aerosol model derived
with the Multi-Wavelength method, scaled by a factor 1000</wcs:description>
            <wcs:name>OMAEROe:AbsorbingAerosolOpticalThicknessMW</wcs:name>
            <wcs:label>Spectral Absorbing Aerosol Optical Thickness for best fit
aerosol model derived with the Multi-Wavelength method, scaled by a factor
1000</wcs:label>
            <wcs:lonLatEnvelope srsName="urn:ogc:def:crs:OGC:1.3:CRS84">
                <gml:pos dimension="2">-179.875000 -89.875000</gml:pos>
                <gml:pos dimension="2">179.875000 89.875000</gml:pos>
            </wcs:lonLatEnvelope>
        </wcs:CoverageOfferingBrief>
        <wcs:CoverageOfferingBrief>
            <wcs:description>Temporally averaged over the requested time for this
product:
Spectral Absorbing Aerosol Optical Thickness for best fit aerosol model derived
with the Multi-Wavelength method, scaled by a factor 1000</wcs:description>
            <wcs:name>OMAEROe:AbsorbingAerosolOpticalThicknessMW_timeAveraged
                </wcs:name>
            <wcs:label>Temporally averaged over the requested time of the Spectral
Absorbing Aerosol Optical Thickness for best fit aerosol model derived with the
Multi-Wavelength method, scaled by a factor 1000</wcs:label>
            <wcs:lonLatEnvelope srsName="urn:ogc:def:crs:OGC:1.3:CRS84">
                <gml:pos dimension="2">-179.875000 -89.875000</gml:pos>
                <gml:pos dimension="2">179.875000 89.875000</gml:pos>
            </wcs:lonLatEnvelope>
        </wcs:CoverageOfferingBrief>
    </wcs:ContentMetadata>
</wcs:WCS_Capabilities>
```


ROA-style services reuse the HTTP operations, and they do not need to define operation on top of HTTP, so that there is no specific section in the service metadata document. They still need to enumerate the list of resources available. This can be done in the service metadata document (this is the case with WMTS 1.0) or might provide a specific resource URL to do it (this is the case with WFS 3.0). Each resource type can be associated to one or more URL templates and the corresponding HTTP supported operations. A URL template is a way to define a pattern for a family of URLs. A URL template looks like a URL, but it contains variable names between "{" and "}", which need to be substituted by values to build a valid URL [10].

An example of a URL template used in WMTS to define all tile URLs of a layer is `http://www.maps.bob/etopo2/default/{TileMatrixSet}/{TileMatrix}/{TileRow}/{TileCol}.png`.

If we substitute {TileMatrixSet} by WholeWorld_CRS_84, {TileMatrix} by 10 m, {TileRow} by 1 and {TileCol} by 3, we get the following valid URL for a particular tile: `http://www.maps.bob/etopo2/default/WholeWorld_CRS_84/10m/1/3.png`.

Recently, in ROA style a new format to describe a web service API emerged: OpenAPI (formerly known as Swagger). OpenAPI is a format that describes the resource types in web services by providing URL templates, the definition of all variables in a URL template, and the format and schema of the expected resource representation. WFS 3.0 is the first OGC standard to incorporate this metadata format, and it is foreseen that others will follow. OpenAPI documents are written in YAML and will be discussed in Sect. 16.6.

16.3.3 Visualization Services

In this section, we describe two OGC services for geospatial data visualization: the Web Map Service (WMS), also published as ISO 19128, and the Web Map Tile Service (WMTS). Both services are mainly intended for map preparation and retrieval. A map is a pictorial representation of geospatial data in an easy to display format (e.g., GIF or JPEG) [11]. It is a combination of geospatial data and a set of symbolization rules. Given an area of a device screen where a remote map needs to be shown, a standard client wanting to use a visualization service has two possible alternatives (Fig. 16.4): to formulate a personalized single WMS GetMap request to cover the exact area of the screen with the exact required extent of the map [12], or to perform a series of WMTS GetTile requests of small tiles that, once displayed one next to the other, will cover the required area of the screen, hiding the exceeded area [13]. After images are positioned on the screen, the user will not see any difference between WMS and WMTS approaches, but when the user tries to navigate quickly through the data, WMTS should outperform WMS due to its internal design. Tiles are optimized for faster delivery since they can be precomputed on the server. Furthermore, the fixed set of tiles enables the use of HTTP network mechanisms for scalability, such as distributed cache systems [14].

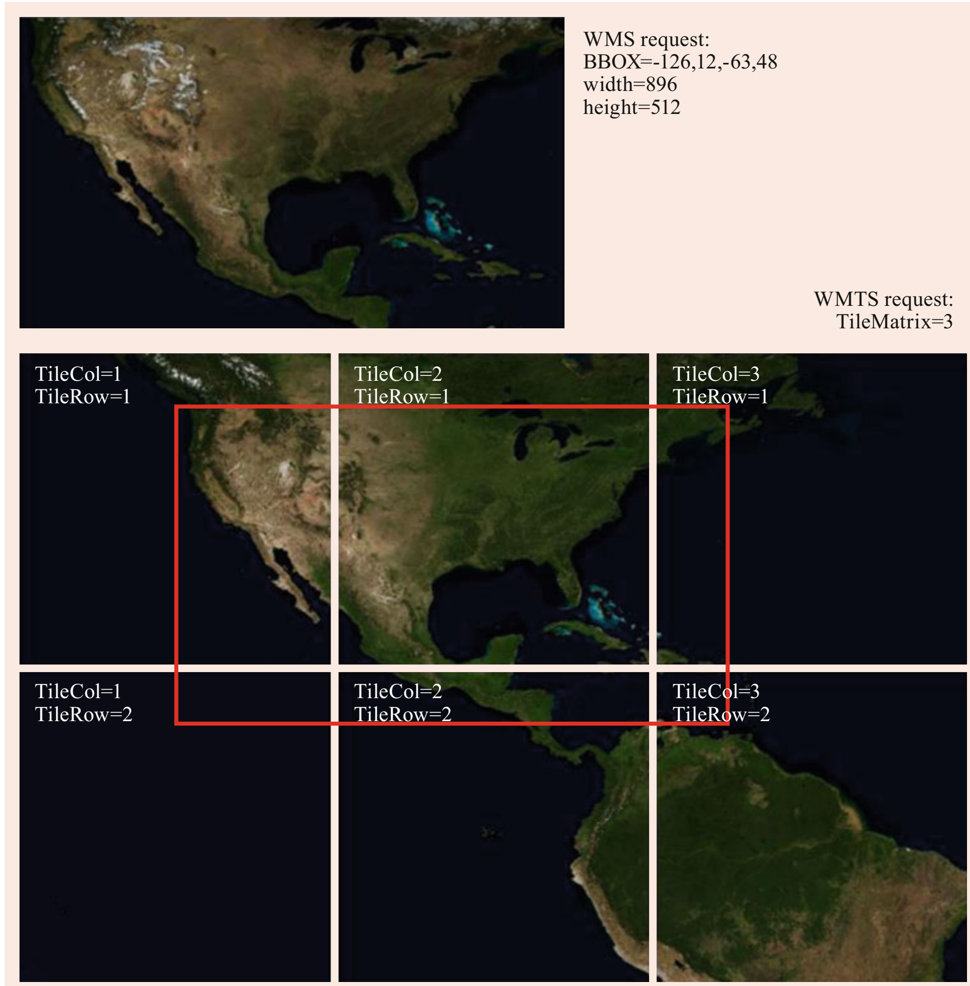


Fig. 16.4 Requesting the same area with a single WMS request and with six WMTS tile requests (courtesy of NASA Earth Observations)

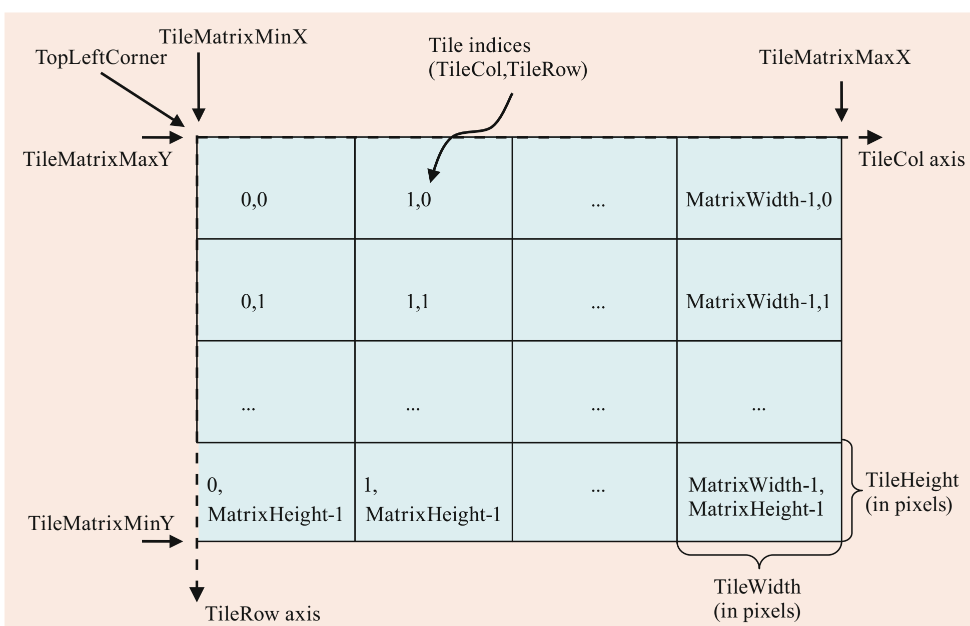


Fig. 16.5 Tiled space

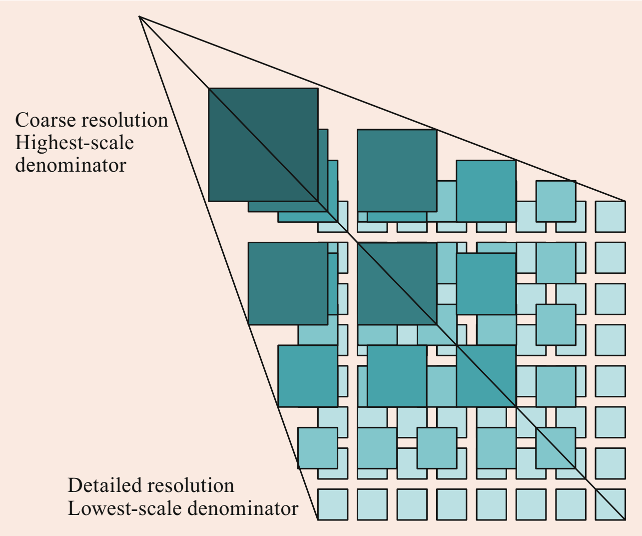


Fig. 16.6 TileMatrixSet representation

In a WMS, a coordinate reference system (CRS) for the data is selected for the two dimensions of the map that are going to be represented defining an implicit linear transformation between the desired bounding box in CRS coordinates (called BBOX) and the two dimensions of the image that starts at 0,0 of the top left corner and ends on the WIDTH,HEIGHT of the lower right corner. WMS can create maps at any desired scale (defined here as the size in the screen divided by the size in the reality). In WMS and WMTS, the scale is calculated by considering that the map image will be shown on a standard 0.28 mm pixel size display. Frequently, the true pixel size is unknown, and 0.28 mm was the actual size of a common display back in 2005. This value is still being used as reference, even if current display devices are built with much smaller pixel sizes.

In WMTS, the width and height are fixed, and the available scales and bounding boxes are discrete and limited to a regular matrix called the TileMatrix (Fig. 16.5). Given an available scale, the size of a pixel can be calculated as $pixelSize = scaleDenominator \times 0.28 \times 10^{-3} / metersPerUnit(crs)$. Tile sizes are $tileWidth \times pixelSize$, $tileHeight \times pixelSpan$. The tiled space starts at a predefined top left corner in CRS coordinates and is filled with tiles one next to one another until MatrixWidth and MatrixHeight are reached [15].

In WMTS, only a list of TileMatrix (each one with its own scale) is allowed. This generally short list of TileMatrix's is known as a TileMatrixSet (Fig. 16.6). A TileMatrix in the TileMatrixSet has a unique alphanumeric identifier in the TileMatrixSet. Some tile-based implementations prefer to use a *zoom level* number, which has the advantage of suggesting some order in the TileMatrixSet. In some other standards, this way of dividing the space is called *image pyramid*-like in clause 11.6 of the OGC KML 2.2 (OGC 07-147r2). JPEG2000 (ISO/IEC 15444-1) and JPIP (ISO/IEC 15444-9) also use a similar division of the space called *resolution levels*. Nevertheless, in those cases, the pyramid is self defined starting from the more detailed TileMatrix (using square tiles), and tiles of the next scales are built by successively aggregating four tiles of the previous scale and then lowering the number pixels by a factor of 2. That approach involves a more rigid structure, which has scale numbers related by powers of 2, and tiles that perfectly overlap four tiles on the immediately inferior scale denominator. TileMatrixSets are more flexible but encompass the more restricted KML *superoverlays* or JPEG2000 resolution levels.

Both WMS and WMTS include three operations (Figs. 16.7 and 16.8), one to get the service metadata (GetCapabilities), one to get a map or a tiled map (GetMap in WMS and GetTile in WMTS), and one to get the actual values represented in a given pixel of a map (GetFeatureInfo) (Tables 16.1 and 16.2). In WMS or WMTS, geospatial information is classified in layers and GetCapabilities enumerates the layer names that can be requested, the possible style names applicable to them, and CRS or TileMatrixSets available. With this information, it is possible to formulate a map or a map tile request using KVP syntax.

A GetFeatureInfo operation has the same parameters as the operation to request maps but adds or overwrites the one shown in Table 16.2.

It is important to note that clients should be able to superimpose map views of information that come from multiple remote and heterogeneous sources into a single view. To help with this process, servers should be able to deliver transparent pixels when there is no data shown for this layer.

WMTS also includes a ROA style to retrieve tiles (and to formulate the equivalent of a GetFeatureInfo query) based on URL templates instead of SOA KVP style. The ServiceMetadata document provides the URL template for each layer.

A WMS can be extended to allow user-defined symbolization of feature or coverage data using the Styled Layer Description (SLD) (Fig. 16.9). This standard adds some extra operations to a WMS to define how the symbology encoding (SE) standard documents can be used with WMS [16]. Before being able to write an SE document, it is necessary to know more about the layers inside the WMS. With the DescribeLayer operation, a user can request information about the features being symbolized (e.g., their feature/coverage type). With this information, the user can define a personalized symbology for the elements present in the layer that can be applied when the GetMap operation is requested. Symbology encoding (SE) specifies an XML format of a map-styling language of the user-defined styling. A GetLegendGraphic operation produces a legend representation of the user-defined symbology.

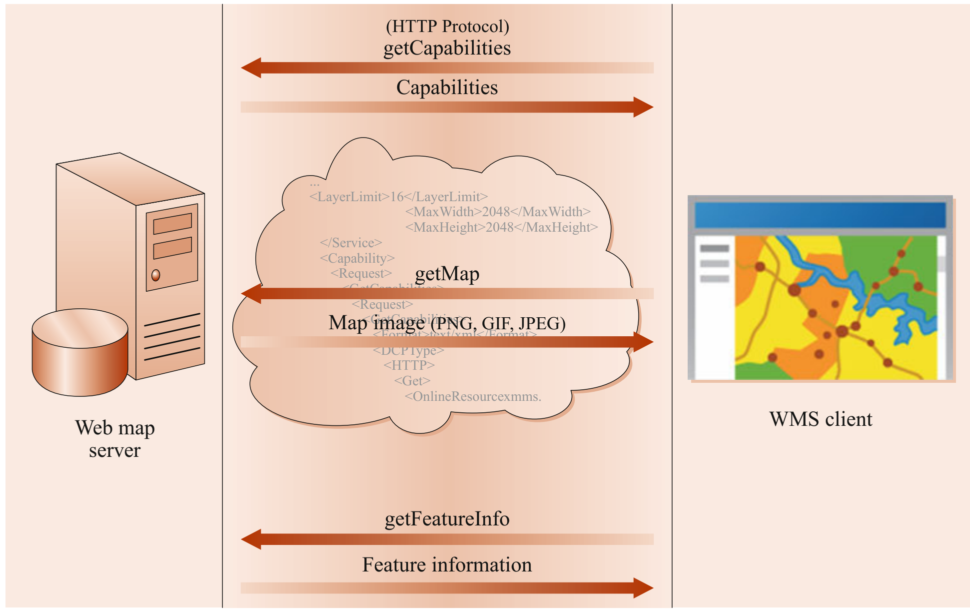


Fig. 16.7 Web Map Service operations

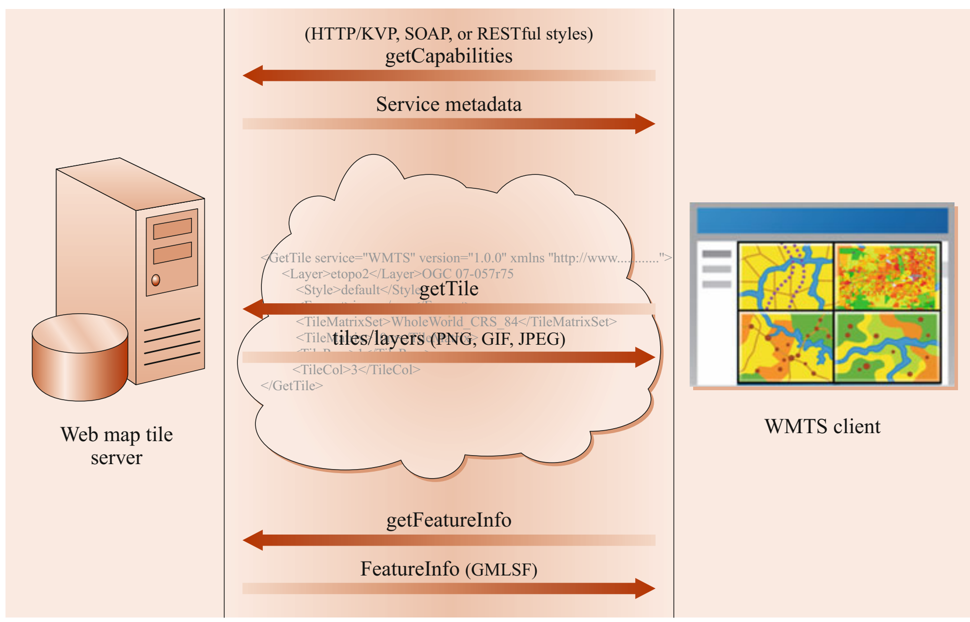


Fig. 16.8 Web map tile service operations

16.3.4 Data Exchange Services

In this section, we will describe three services for accessing, retrieving and, in some cases, uploading geospatial information in a distributed environment that targets three different data models. The web feature service (WFS) exchanges geospatial information in the vector data models (called "features"), the web coverage service (WCS) retrieves geospatial information in the raster model (called "coverages"), and the sensor observation service (SOS) retrieves geospatial information in the sensor models (called "Observations and Measurements", O&M) (Table 16.3). Since the three services follow the OWS Common model, in a first approach, we can describe them together. The three services provide an operation to retrieve the service metadata. This document enumerates the names of the resources available in the service. With these names, we can get more information about them with a "describe" operation. Finally, we can retrieve the actual resources in their representation: WFS delivers the features in GML, WCS delivers the coverages mainly in raster formats (e.g., GeoTIFF, NetCDF …), and SOS delivers the observations in O&M XML encoding. Sometimes, there is the need to limit a resource collection to a fragment of it by using some filtering capability

Table 16.1 GetMap and GetTile request parameters

WMS		WMTS	
Parameter	Meaning	Parameter	Meaning
Service	`WMS`	**Service**	`WMTS`
Request	`GetMap`	**Request**	`GetTile`
Version	`1.3.0`	**Version**	`1.0.0`
Layers	List of layer names	**Layer**	Layer name
Styles	A style for each layer name	**Style**	Style name
Format	Output format	**Format**	Output format
CRS	CRS identifier	**TileMatrixSet**	TileMatrixSet identifier
BBox	Bounding box 4 coordinates	**TileMatrix**	TileMatrix identifier
		TileRow	Row index of the tile
		TileCol	Column index of the tile
Width	Map width		
Height	Map height		
Transparent	`TRUE` or `FALSE`		
Other dimension name	Other dimension value	***Other dimension name***	Other dimension value

Table 16.2 GetFeatureInfo request parameters

WMS		WMTS	
Parameter	Meaning	Parameter	Meaning
Service	`WMS`	**Service**	`WMTS`
Request	`GetFeatureInfo`	**Request**	`GetFeatureInfo`
Version	`1.3.0`	**Version**	`1.0.0`
InfoFormat	Output format	**InfoFormat**	Output format
I	Row index of a pixel within the map	**I**	Row index of a pixel within the tile
J	Column index of a pixel within the map	**J**	Column index of a pixel within the tile

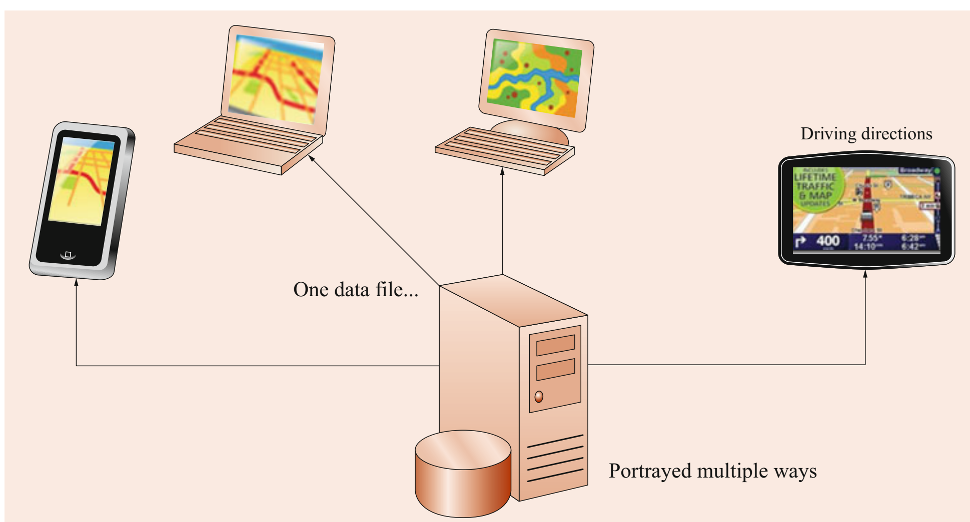


Fig. 16.9 Style Layer Descriptor

Table 16.3 Data exchange services functionalities

Functionality	WFS 2.0	WCS 2.0	SOS 2.0
Service metadata	GetCapabilities	GetCapabilities	GetCapabilities
Resources	Feature type names	Coverage names	Observation offerings
Resource description	DescribeFeatureType	DescribeCoverage	DescribeSensor
Retrieve resources	GetFeature	GetCoverage	GetFeatureOfInterest
Retrieve parts of resources	GetValueProperty	GetCoverage	GetObservation
Create, update, or delete resources	Transaction	N/A	InsertSensor, DeleteSensor, InsertObservation

Fig. 16.10 USGS airport runway GML applications schema (DescribeFeatureType; simplified)

```
<?xml version="1.0" encoding="utf-8" ?>
<xs:schema>
  <xs:element name='Airport_Runway' type='WFS_transportation:Airport_RunwayType'
substitutionGroup='gml:AbstractFeature'/>
    <xs:complexType name='Airport_RunwayType'>
      <xs:complexContent>
        <xs:extension base='gml:AbstractFeatureType'>
          <xs:sequence>
            <xs:element name='OBJECTID' type='xs:int'/>
            <xs:element minOccurs='0' name='Shape'
type='gml:MultiSurfacePropertyType'/>
            <xs:element minOccurs='0' name='PERMANENT_IDENTIFIER'>
                <xs:simpleType>
                      <xs:restriction base='xs:string'>
                            <xs:maxLength value='40'/>
                      </xs:restriction>
                </xs:simpleType>
            </xs:element>
            <xs:element minOccurs='0' name='SOURCE_FEATUREID'>
                <xs:simpleType>...</xs:simpleType>
            </xs:element>
            <xs:element minOccurs='0' name='SOURCE_DATASETID'>
                <xs:simpleType>...</xs:simpleType>
            </xs:element>
            <xs:element minOccurs='0' name='SOURCE_DATADESC'>
                <xs:simpleType>...</xs:simpleType>
            </xs:element>
            <xs:element minOccurs='0' name='SOURCE_ORIGINATOR'>
                <xs:simpleType>...</xs:simpleType>
            </xs:element>
            <xs:element minOccurs='0' name='Data_Security' type='xs:short'/>
            <xs:element minOccurs='0' name='Distribution_Policy'>
                <xs:simpleType>...</xs:simpleType>
            </xs:element>
            <xs:element minOccurs='0' name='LOADDATE' type='xs:dateTime'/>
            <xs:element minOccurs='0' name='FCode' type='xs:int'/>
            <xs:element minOccurs='0' name='RUNWAY_ID'>
                <xs:simpleType>...</xs:simpleType>
            </xs:element><xs:element minOccurs='0' name='FAA_AIRPORT_CODE'>
                <xs:simpleType>...</xs:simpleType>
            </xs:element>
            <xs:element minOccurs='0' name='CITY'>
                <xs:simpleType>...</xs:simpleType>
            </xs:element>
            <xs:element minOccurs='0' name='STATE'>
                <xs:simpleType>...</xs:simpleType>
            </xs:element>
            <xs:element minOccurs='0' name='OwnerType' type='xs:int'/>
            <xs:element minOccurs='0' name='UseStatus' type='xs:int'/>
            <xs:element minOccurs='0' name='SHAPE_Length' type='xs:double'/>
            <xs:element minOccurs='0' name='SHAPE_Area' type='xs:double'/>
          </xs:sequence>
        </xs:extension>
      </xs:complexContent>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

in the retrieve operation or by using other similar operations. Finally, WFS and SOS define transactional operations that allow creating, updating, or deleting resources on the server.

To know more about these data services there is a need to know more about the underlying data model of the resources they use. This is the purpose of the following sections.

GML and Web Feature Service

WFS [17] uses the general feature model (GFM) and the GML encoding to represent the geospatial features. To understand WFS, first you should understand GML. Unfortunately, understanding GML is not an easy task, in our opinion, due to two main reasons: the flexibility of the GFM, combined with the complexities of XML schemas. To rep-

Fig. 16.11 USGS airport runway encoded in GML (GetFeature response; simplified)

```
<wfs:FeatureCollection>
  <gml:member>
    <Airport_Runway gml:id='F10__13355'>...</Airport_Runway>
  </gml:member>
  <gml:member>
    <Airport_Runway gml:id='F10__13507'>...</Airport_Runway>
  </gml:member>
  <gml:member>
    <Airport_Runway gml:id='F10__13526'>
      <OBJECTID>13526</OBJECTID>
      <Shape>
        <gml:MultiSurface srsName="EPSG:3857">
          <gml:surfaceMember>
            <gml:Polygon>
                <gml:exterior>
                      <gml:LinearRing>
                            <gml:posList> -8575942.8157000002
4701809.001000002 -8575720.8753999993 4699005.684799999 -8575664.9667000007
4699010.873800002 -8575884.6827000007 4701813.5922999978 -8575942.8157000002
4701809.001000002</gml:posList>
                      </gml:LinearRing>
                </gml:exterior>
            </gml:Polygon>
          </gml:surfaceMember>
        </gml:MultiSurface>
      </Shape>
      <PERMANENT_IDENTIFIER>f4156a1d-9259-46d8-abe1-
c68b689e2c12</PERMANENT_IDENTIFIER>
      <SOURCE_DATASETID>{49981A0D-FC48-4BBF-872D-
8E72CE83EA8B}</SOURCE_DATASETID>
      <SOURCE_DATADESC>Public and Private Airports</SOURCE_DATADESC>
      <SOURCE_ORIGINATOR>Federal Aviation Administration</SOURCE_ORIGINATOR>
      <Data_Security>5</Data_Security>
      <Distribution_Policy>E4</Distribution_Policy>
      <LOADDATE>2016-06-06T14:16:43</LOADDATE>
      <FCode>20100</FCode>
      <RUNWAY_ID>01/19</RUNWAY_ID>
      <FAA_AIRPORT_CODE>DCA</FAA_AIRPORT_CODE>
      <CITY>Washington</CITY>
      <STATE>VA</STATE>
      <OwnerType>1</OwnerType>
      <UseStatus>1</UseStatus>
      <SHAPE_Length>5737.8685680632716</SHAPE_Length>
      <SHAPE_Area>160910.31463661761</SHAPE_Area>
    </Airport_Runway>
  </gml:member>
</wfs:FeatureCollection>
```

resent features in GML, first we need a data model and to encode it into a GML applications schema [18]. In practice, this means that features are arranged in feature types. Each feature type will receive a name and will define the list of properties that this feature type has. Let us take the USGS WFS service for transportation data. The GetCapabilities request has a list of feature types:

Request: `https://services.nationalmap.gov/arcgis/services/WFS/transportation/MapServer/WFSServer?request=GetCapabilities&service=WFS`

List of feature types: `Airport_Runway`, `Airport`, `County_Route`, `Interstate`, `Local_Road`, `State_Route`, `US_Railroad`, and `US_Route`.

To obtain the GML application schema for the feature type `Airport_Runway`, we can execute a DescribeFeatureType request [19]: `https://services.nationalmap.gov/arcgis/services/WFS/transportation/MapServer/WFSServer?request=DescribeFeatureType&service=WFS&typename=WFS_transportation:Airport_Runway` (Fig. 16.10).

In the response (a GML application schema), the feature type tells us that all "runway" features have the same properties and what they are. The most relevant ones are GML MultiSurface geometry (`Shape`), the FAA name of the airport (`FAA_AIRPORT_CODE`), the city (`CITY`), and the runway name (`RUNWAY_ID`). Now we know what we

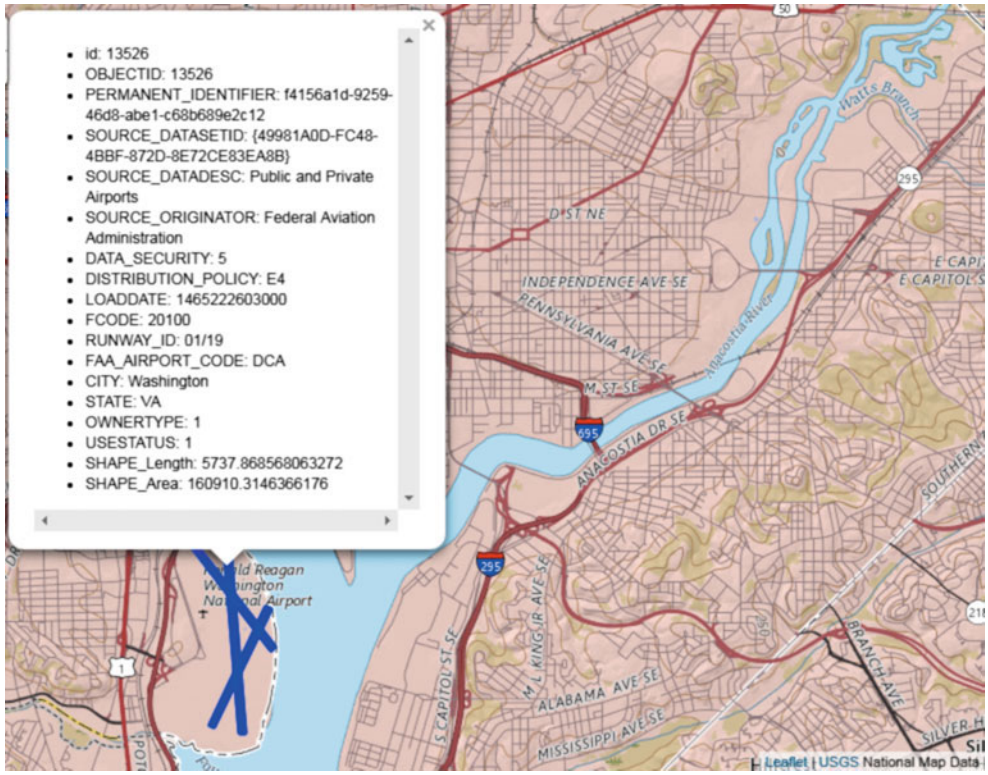


Fig. 16.12 USGS airport runway (previously seen as GML in Fig. 16.11) seen in the Leaflet WFS test browser (courtesy of USGS National Map Data)

can expect from the data retrieved and, eventually, we can use this information in a GetValueProperty request.

The three runways of the DCA Washington (Ronald Reagan National Airport) can be retrieved with a GetFeature request limited by airport runway feature type name and bounding box (Fig. 16.11).

Request: `https://services.nationalmap.gov/arcgis/services/WFS/transportation/MapServer/WFSServer?request=GetFeature&service=WFS&typename=WFS_transportation:Airport_Runway&bbox=-77.092,38.842,-76.992,38.942`

WFS services offer geospatial data that can easily be big. Formulating requests that try to retrieve the complete dataset are normally rejected by services after some minutes of processing. To prevent that it is very important to know how to limit the GetFeature requests by a bounding box (or other restrictive criteria) in order to aim for results that are small enough for the server to handle. In the previous example, we provided a bounding box that limits the request to the Washington DC area. We see that we got three features, one for each runway at the DCA airport. The longest runway is the `01/19` with 5.7 km.

The GML geometrical features can be shown in a feature viewer with the other properties of the north–south runway (Fig. 16.12).

The three operations presented here are complemented by others such as the transactional ones (Fig. 16.13).

In order to simplify the usage of WFS, OGC introduced the OGC API – Features (formerly known as WFS 3.0) which removes the need for a GML application schema by adopting GeoJSON as a default exchange format. More details about this standard will be provided in Sect. 16.6.

CIS and Web Coverage Service

Commonly, GML is associated to vector data, but actually Ver. 3 introduced raster support. Again, GML requires associating a GML application schema to each raster described in GML. This is rather tedious and repetitive because most raster documents are alike. WCS wanted to adopt GML to describe coverage data but in order to simplify the use of WCS, WCS avoided the need to create a GML application schema by defining a common GML application schema that is valid for all coverages. It is called the Coverage Implementation Schema (CIS; formerly known as GMLCov) [20]. Following the CIS model, a coverage can be defined by describing three aspects: DomainSet, RangeType, and RangeSet (Fig. 16.14).

In the DomainSet, the axes of the multidimensional grid space are defined. In a regular referenced grid, it provides the names of the axes, as well as the origin and the spacing of each axis. In a 2-D space, it defines two axes that create a regular 2-D grid of cells that covers some space, but the cells are still empty and have no values in them. The rangeType defines the types of values (e.g., float numbers), the units, and the nodata values that are going to populate the grid. In the rangeSet, the actual list of values that populate the cells

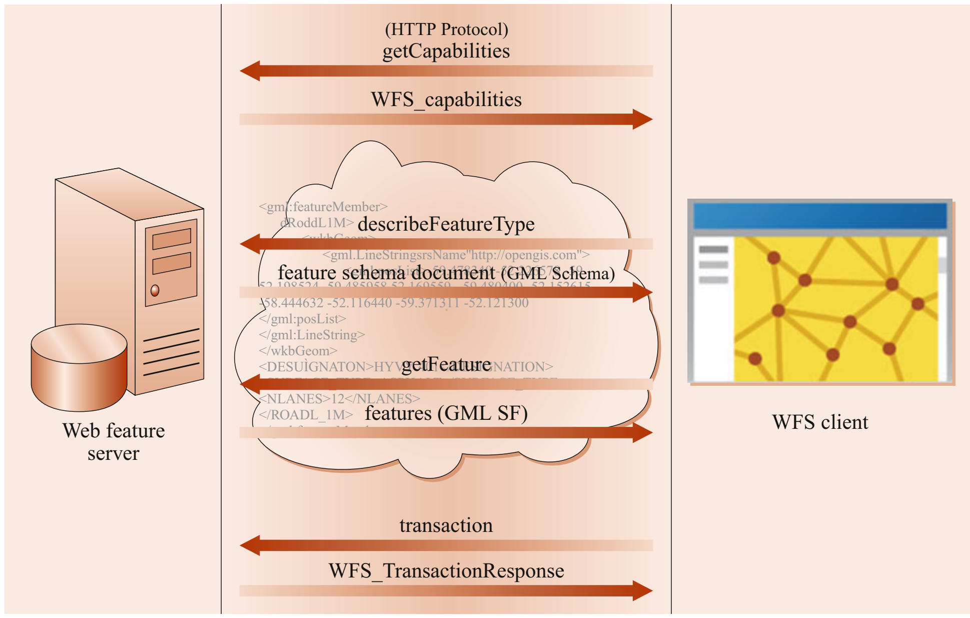


Fig. 16.13 Web Feature Service operations

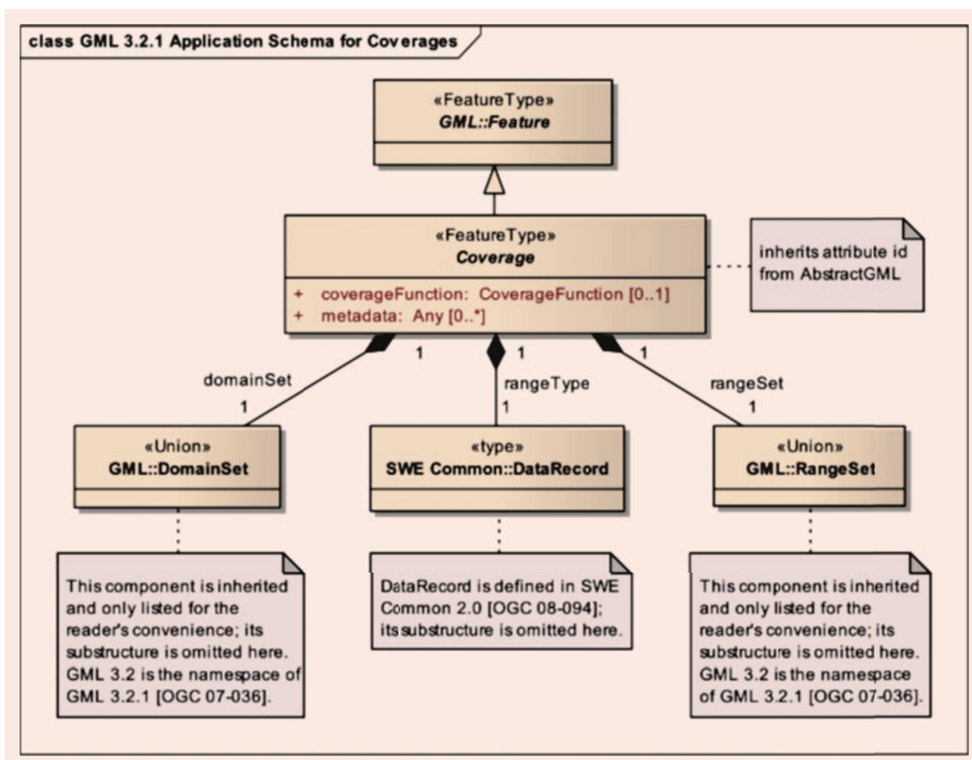


Fig. 16.14 The coverage Implementation Schema (fragment) (courtesy of OGC)

is given (e.g., in a Digital Terrain Model (DTM) it will be a long list of elevations). The DomainSet and the rangeType are normally encoded in XML. The rangeSet could also be included in the XML but is commonly referenced to an external raster file such as a GeoTIFF or a JPEG2000 for a more compact notation. This data model has its origins in GML elements that were extracted to define the common application schema for coverages.

In a WCS, information is structured in coverages [21]. A server may have one or more coverages to offer. Let us take the Rasdaman EarthServer demo service. The GetCapabilities request has a list of coverages:

Request: `http://ows.rasdaman.org/rasdaman/ows?&SERVICE=WCS&ACCEPTVERSIONS=2.0.1&REQUEST=GetCapabilities`

Fig. 16.15 German DTM domainSet and the rangeType (DescribeCoverage response; simplified)

```xml
<?xml version="1.0" encoding="UTF-8"?>
<wcs:CoverageDescriptions>
    <wcs:CoverageDescription gml:id="Germany_DTM">
        <boundedBy>...</boundedBy>
        <wcs:CoverageId>Germany_DTM</wcs:CoverageId>
        <coverageFunction>...</coverageFunction>
        <domainSet>
            <RectifiedGrid dimension="2" gml:id="Germany_DTM-grid">
                <limits>
                    <GridEnvelope>
                        <low>0 0</low>
                        <high>10800 13200</high>
                    </GridEnvelope>
                </limits>
                <axisLabels>Lat Long</axisLabels>
                <origin>
                    <Point gml:id="Germany_DTM-origin"
srsName="http://ows.rasdaman.org/def/crs/EPSG/0/4326">
                        <pos>56.0 5.0</pos>
                    </Point>
                </origin>
                <offsetVector
srsName="http://ows.rasdaman.org/def/crs/EPSG/0/4326">-0.00083333 0</offsetVector>
                <offsetVector
srsName="http://ows.rasdaman.org/def/crs/EPSG/0/4326">0 0.000833333</offsetVector>
            </RectifiedGrid>
        </domainSet>
        <gmlcov:rangeType>
            <swe:DataRecord>
                <swe:field name="Elevation">
                    <swe:Quantity>
                        <swe:label>Elevation</swe:label>
                        <swe:uom code="m"/>
                    </swe:Quantity>
                </swe:field>
            </swe:DataRecord>
        </gmlcov:rangeType>
        <wcs:ServiceParameters>
            <wcs:CoverageSubtype>RectifiedGridCoverage</wcs:CoverageSubtype>
            <wcs:nativeFormat>application/octet-stream</wcs:nativeFormat>
        </wcs:ServiceParameters>
    </wcs:CoverageDescription>
</wcs:CoverageDescriptions>
```

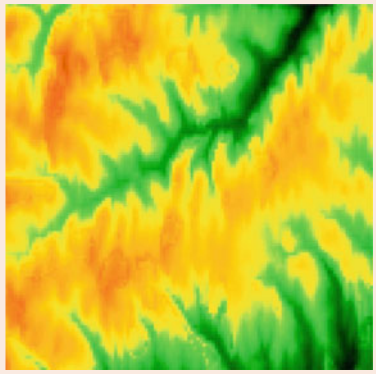
List of coverage identifiers:

- `AverageChloroColorScaled`,
- `AverageChlorophyllScaled`,
- `AvgLandTemp`,
- `AvgTemperatureColorScaled`,
- `BlueMarbleCov`,
- `climate_cloud`,
- `climate_earth`,
- `Germany_DTM`,
- `lena`,
- `mean_summer_airtemp`,
- `meris_lai, multiband`,
- `NIR`,
- `NN3_1`, `NN3_2`, `NN3_3`, `NN3_4`,
- `RadianceColor`,
- `RadianceColorScaled`,
- `Temperature4D`,
- `test_irr_cube_2`,
- `test_mean_summer_airtemp`,
- `visible_human`.

To obtain the domainSet and the rangeType of the CIS schema for the `Germany_DTM`, we can execute a DescribeCoverage request (Fig. 16.15): `http://ows.rasdaman.org/rasdaman/ows?&SERVICE=WCS&VERSION=2.0.1&REQUEST=DescribeCoverage&COVERAGEID=Germany_DTM`

In the description, we can see the definition of a 2-D grid of $10{,}800 \times 13{,}200$ cells, representing latitude and longitude that starts at 56.0, 5.0 and has a cell size of 0.00083333 degrees. The cells will be populated with elevations.

The central part of the German DTM can be retrieved with a GetCoverage request indicating the coverage name and limited by a bounding box (Fig. 16.16).



```
<boundedBy>
  <Envelope axisLabels="Lat Long" srsDimension="2">
    <lowerCorner>49.99958333 9.99958333</lowerCorner>
     <upperCorner>50.10041667 10.10041667</upperCorner>
        </Envelope>
    </boundedBy>
    <domainSet>
      <RectifiedGrid dimension="2">
        <limits>
           <GridEnvelope>
             <low>7080 6000</low>
             <high>7200 6120</high>
           </GridEnvelope>
        </limits>
        <axisLabels>Lat Long</axisLabels>
...
```

Fig. 16.16 German DTM fragment shown as a raster and also in a CIS fragment (Courtesy of OCLC)

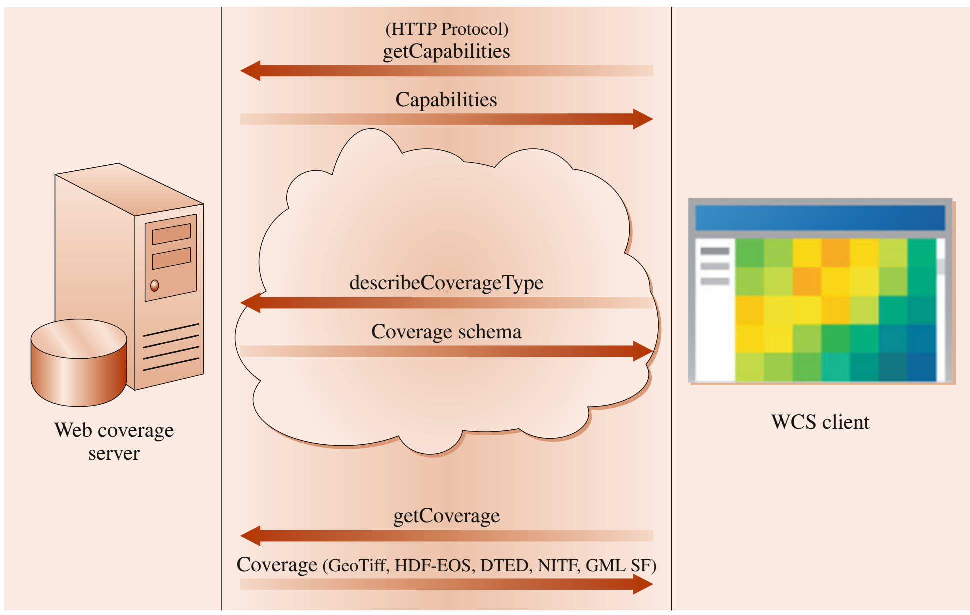


Fig. 16.17 Web Coverage Service operations

Request: `http://ows.rasdaman.org/rasdaman/ows?&SERVICE=WCS&VERSION=2.0.1&REQUEST=GetCoverage&COVERAGEID=Germany_DTM&SUBSET=Lat(50,50.1)&SUBSET=Long(10,10.1)&FORMAT=image/tiff`

Figure 16.17 represents the three most popular operations in WCS. Again, WCS services are commonly offering geospatial data that can easily be big. Requests that try to retrieve the complete dataset are normally rejected by services after some minutes of processing. To prevent this, it is very important to know how to limit the GetCoverage requests by a bounding box (or other restrictive criteria) in order to anticipate a result that is small enough for the server to handle [22]. In this case, we use the subsetting extension of WCS 2.0 to request a subset for each axis.

The result of this operation will be a small tiff file with the elevations of the requested area. Changing `FORMAT=image/tiff` to `FORMAT=application/gml+xml` in the GetCoverage request will result in a full CIS document describing the domainSet, the rangeType, and providing a long list of integer numbers as the rangeSet.

O&M and the Sensor Observation Service

The observations and measurements (O&M) data model (also published as ISO 19156) is the basis for the sensor observation service (SOS) standard [23]. O&M defines a conceptual schema for observations, and for features involved when making observations. An observation is an act at a discrete instant or period, through which a number or term is assigned to a phenomenon using a procedure, such as a sensor, instrument, or algorithm. Observations commonly involve sampling of an ultimate feature of interest. Figure 16.18 summarizes the different aspects associated with the O&M model and compares them with the basic UML model for the observations. An observation is divided into the actual FeatureOfInterest that will contain the geospatial

16

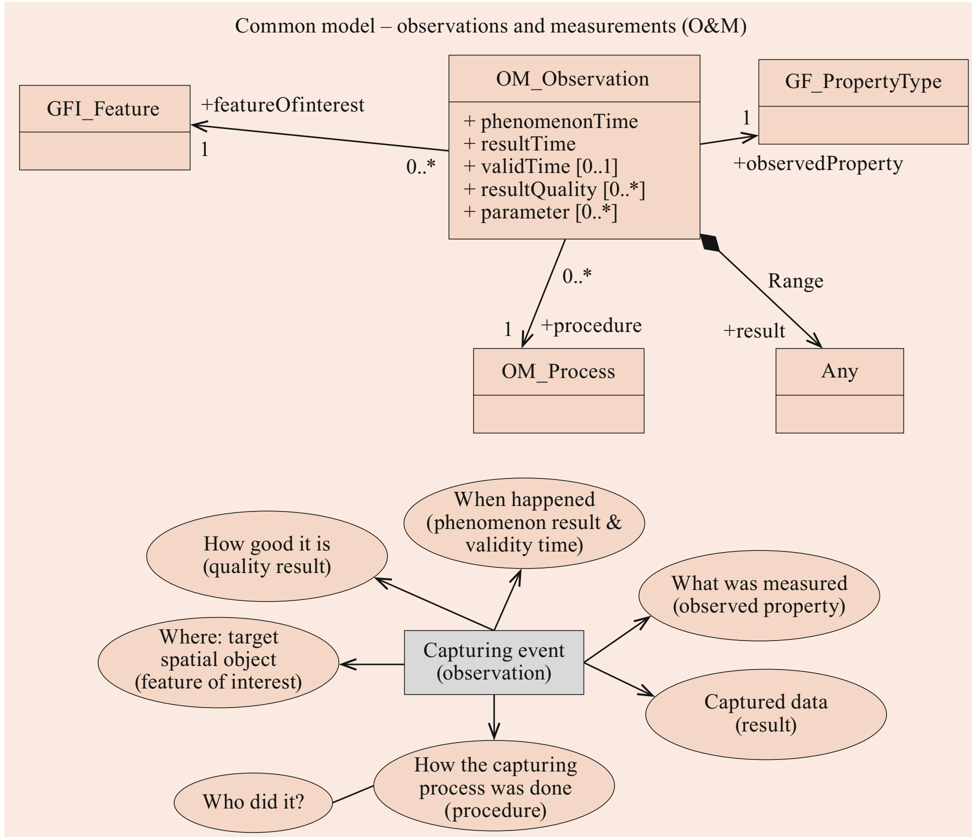


Fig. 16.18 O&M basic model

aspect, the observed property that will contain the semantical definition of what is being observed, the process used to do the measurement (that includes the sensor and the responsible person behind), when the measure was done and archived, and the actual result of the measurement that needs to be flexible to accommodate any kind of sensor output, including isolated values (a temperature on a thermometer), a time series (a sequence of temperatures) an image (a temperature spatial distribution), a video (the evolution of a temperature distribution), a citizen science measurement (the perception of the temperature), etc.

The way the SOS service is used differs a bit from the other services described so far. In this case, the service metadata that you get from a GetCapabilities returns the content grouped into ObservationOfferings. An ObservationOffering groups collections of observations produced by the same procedure (e.g., the same sensor system) [24]. The ObservationOffering lists the basic metadata for the associated observations, including the observed properties of the observations. Be aware that an observation may belong to more than one ObservationOffering. An ObservationOffering is not directly related to observation collections in the same hierarchical structure that features are grouped in FeatureTypes, or cells were grouped into coverages, so ObservationOffering cannot be directly used as a filter parameter for requesting observations. It is useful to know about the observation procedure (the sensors used) and to request more information filtering by procedure identifies [25]. The service metadata can be obtained with a request like this:

Request: `http://sensorweb.demo.52north.org/52n-sos-webapp/service?service=SOS&request=GetCapabilities`

Figure 16.19 shows the structure the service metadata and the details of an ObservationOffering that refers to the procedure identifier `http://www.52north.org/test/procedure/1`

From the ObservationOffering, we can extract the procedure identifiers and request more details about them with DescribeSensor.

Request: `http://sensorweb.demo.52north.org/52n-sos-webapp/service?service=SOS&version=2.0.0&request=DescribeSensor&procedure=http%3A%2F%2Fwww.52north.org%2Ftest%2Fprocedure%2F1&procedureDescriptionFormat=http%3A%2F%2Fwww.opengis.net%2FsensorML%2F1.0.1`

The return is a very detailed description of the sensor based on another standards from the sensor web enablement (SWE) family: SensorML. The description of this standard is out of scope of this chapter.

We can also request information about where the observation took place by requesting the geometrical information included in the FeatureOfInterest part of the O&M model.

Fig. 16.19 Service metadata document fragment (GetCapabilities response; simplified)

```xml
<?xml version="1.0" encoding="UTF-8"?>
<sos:Capabilities>
  <ows:ServiceIdentification>...</ows:ServiceIdentification>
  <ows:ServiceProvider>...</ows:ServiceProvider>
  <ows:OperationsMetadata>...</ows:OperationsMetadata>
  <sos:filterCapabilities>...</sos:filterCapabilities>
  <sos:contents>
    <sos:Contents>
      ...
      <swes:offering>
        <sos:ObservationOffering xmlns:ns="http://www.opengis.net/sos/2.0">

<swes:identifier>http://www.52north.org/test/offering/1</swes:identifier>
          <swes:name codeSpace="eng">Offering for sensor 1</swes:name>
          <swes:procedure>http://www.52north.org/test/procedure/1</swes:procedure>
          <swes:procedureDescriptionFormat>http://www.opengis.net/sensorML/1.0.1
          </swes:procedureDescriptionFormat>
          <swes:procedureDescriptionFormat>http://www.opengis.net/sensorml/2.0
          </swes:procedureDescriptionFormat>
          <swes:observableProperty>http://www.52north.org/test/
          observableProperty/1</swes:observableProperty>
          <swes:observableProperty>http://www.52north.org/test/
          observableProperty/9_1</swes:observableProperty>
          <swes:observableProperty>http://www.52north.org/test/
          observableProperty/9_2</swes:observableProperty>
          <sos:observedArea>
            <gml:Envelope srsName="http://www.opengis.net/def/crs/EPSG/0/4326">
              <gml:lowerCorner>51.883906 7.727958</gml:lowerCorner>
              <gml:upperCorner>51.883906 7.727958</gml:upperCorner>
            </gml:Envelope>
          </sos:observedArea>
          <sos:phenomenonTime>
            <gml:TimePeriod gml:id="phenomenonTime_4">
              <gml:beginPosition>2012-11-19T13:00:00.000Z</gml:beginPosition>
              <gml:endPosition>2012-11-19T13:09:00.000Z</gml:endPosition>
            </gml:TimePeriod>
          </sos:phenomenonTime>
          <sos:resultTime>
            <gml:TimePeriod gml:id="resultTime_4">
              <gml:beginPosition>2012-11-19T13:00:00.000Z</gml:beginPosition>
              <gml:endPosition>2012-11-19T13:09:00.000Z</gml:endPosition>
            </gml:TimePeriod>
          </sos:resultTime>
          <sos:responseFormat>application/json</sos:responseFormat>
          <sos:observationType>http://www.opengis.net/def/observationType/OGC-
OM/2.0/OM_Measurement</sos:observationType>

<sos:featureOfInterestType>http://www.opengis.net/def/samplingFeatureType/OGC-
OM/2.0/SF_SamplingPoint</sos:featureOfInterestType>
        </sos:ObservationOffering>
      </swes:offering>
      ...
    </sos:Contents>
  </sos:contents>
</sos:Capabilities>
```

We can request all FeaturesOfInterest or filter them by some criteria; for example, the procedure (Fig. 16.20):

Request: `http://sensorweb.demo.52north.org/52n-sos-webapp/service?service=SOS&version=2.0.0&request=GetFeatureOfInterest&procedure=http%3A%2F%2Fwww.52north.org%2Ftest%2Fprocedure%2F1`

In the simplest case, we will get the position of the sensor as a point.

Finally, each feature reported in the GetFeatureOfInterest response presents an identifier that can be used to request the actual observations done in this position.

Request: `http://sensorweb.demo.52north.org/52n-sos-webapp/service?service=SOS&version=2.0.0&request=GetObservation&featureOfInterest=http%3A%2F%2Fwww.52north.org%2Ftest%2FfeatureOfInterest%2F1`

Fig. 16.20 GetFeatureOfInterest document response (simplified)

```
<?xml version="1.0" encoding="UTF-8"?>
<sos:GetFeatureOfInterestResponse>
  <sos:featureMember>
    <sams:SF_SpatialSamplingFeature
gml:id="ssf_4242C39825B3361F8D6B329C3B95AABBC0AE7190">
      <gml:identifier
codeSpace="http://www.opengis.net/def/nil/OGC/0/unknown">http://www.52north.org/te
st/featureOfInterest/1</gml:identifier>
      <gml:name>con terra</gml:name>
      <sf:type xlink:href="http://www.opengis.net/def/samplingFeatureType/OGC-
OM/2.0/SF_SamplingPoint"/>
      <sf:sampledFeature
xlink:href="http://www.52north.org/test/featureOfInterest/world"/>
      <sams:shape>
        <gml:Point gml:id="point_ssf_4242C39825B3361F8D6B329C3B95AABBC0AE7190">
          <gml:pos srsName="http://www.opengis.net/def/crs/EPSG/0/4326">51.883906
7.727958</gml:pos>
        </gml:Point>
      </sams:shape>
    </sams:SF_SpatialSamplingFeature>
  </sos:featureMember>
</sos:GetFeatureOfInterestResponse>
```

Fig. 16.21 GetObservation document response (simplified)

```
<?xml version="1.0" encoding="UTF-8"?>
<sos:GetObservationResponse>
  <sos:observationData>
    <om:OM_Observation gml:id="o_121">
      <gml:identifier>http://www.52north.org/test/observation/1</gml:identifier>
      <om:type xlink:href="http://www.opengis.net/def/observationType/OGC-
OM/2.0/OM_Measurement"/>
      <om:phenomenonTime>
        <gml:TimeInstant gml:id="phenomenonTime_121">
          <gml:timePosition>2012-11-19T13:00:00.000Z</gml:timePosition>
        </gml:TimeInstant>
      </om:phenomenonTime>
      <om:resultTime xlink:href="#phenomenonTime_121"/>
      <om:procedure xlink:href="http://www.52north.org/test/procedure/1"/>
      <om:observedProperty
xlink:href="http://www.52north.org/test/observableProperty/1"/>
      <om:featureOfInterest
xlink:href="http://www.52north.org/test/featureOfInterest/1" xlink:title="con
terra"/>
      <om:result xmlns:ns="http://www.opengis.net/gml/3.2" uom="test_unit_1"
xsi:type="ns:MeasureType">1.0</om:result>
    </om:OM_Observation>
  </sos:observationData>
  <sos:observationData>
    <om:OM_Observation gml:id="o_122">
      <gml:identifier
codeSpace="http://www.opengis.net/def/nil/OGC/0/unknown">http://www.52north.org/te
st/observation/2</gml:identifier>
      <om:type xlink:href="http://www.opengis.net/def/observationType/OGC-
OM/2.0/OM_Measurement"/>
      <om:phenomenonTime>
        <gml:TimeInstant gml:id="phenomenonTime_122">
          <gml:timePosition>2012-11-19T13:01:00.000Z</gml:timePosition>
        </gml:TimeInstant>
      </om:phenomenonTime>
      <om:resultTime xlink:href="#phenomenonTime_122"/>
      <om:procedure xlink:href="http://www.52north.org/test/procedure/1"/>
      <om:observedProperty
xlink:href="http://www.52north.org/test/observableProperty/1"/>
      <om:featureOfInterest
xlink:href="http://www.52north.org/test/featureOfInterest/1" xlink:title="con
terra"/>
      <om:result xmlns:ns="http://www.opengis.net/gml/3.2" uom="test_unit_1"
xsi:type="ns:MeasureType">1.1</om:result>
    </om:OM_Observation>
  </sos:observationData>
</sos:GetObservationResponse>
```

Fig. 16.22 WPS service metadata document response (fragment)

```
<?xml version="1.0" encoding="UTF-8"?>
<wps:Capabilities service="WPS" version="2.0.0">
  <ows:ServiceIdentification>...</ows:ServiceIdentification>
  <ows:ServiceProvider>...</ows:ServiceProvider>
  <ows:OperationsMetadata>...</ows:OperationsMetadata>
  <wps:Contents>
    ...
    <wps:ProcessSummary processVersion="1.1.0" jobControlOptions="sync-execute
async-execute" outputTransmission="value reference">
      <ows:Title>org.n52.wps.server.algorithm.SimpleBufferAlgorithm</ows:Title>
      <ows:Identifier>org.n52.wps.server.algorithm.SimpleBufferAlgorithm
      </ows:Identifier>
      <ows:Metadata xlink:role="Process description"
xlink:href="http://geoprocessing.demo.52north.org/wps/WebProcessingService?service
=WPS&amp;request=DescribeProcess&amp;version=2.0.0&amp;identifier=org.n52.wps.serv
er.algorithm.SimpleBufferAlgorithm"/>
    </wps:ProcessSummary>
    ...
  </wps:Contents>
</wps:Capabilities>
```

In Fig. 16.21, we see two observations done at to instants in time separated by 1 min, and we can note that the result of the measure has increased from 1.0 to 1.1.

There are some other operations available that will not be described here but can be found in the corresponding standard document.

16.3.5 Processing Services

OGC has produced two standards for processing services: the web processing service (WPS) and the web coverage processing service (WCPS). The WPS is a flexible service that can offer any kind of processing on the web. In contrast, WCPS offers a predefined list of operation for coverages. WCPS is not described in this chapter.

WPS defines an interface that facilitates the publishing of geospatial processes. Processes include any end-to-end algorithm, calculation, or model that operates on spatially referenced data that will not require any intermediate interaction with a client or another process. A WPS may offer calculations as simple as subtracting two coverages of the same variable (e.g., determining the difference between two temperature models), or as complex as a global carbon model. A client can execute a process by providing information about the input parameters of the process, which may include data that is known by the WPS, URLs to resources to download before the process start, or access requests to OGC web services such as WFS, WCS, or SOS. A client can execute an atomic calculation on a single WPS or can orchestrate a chaining of WPS processes combining several processing and access services [26]. A client can store the chain and create a repeatable workflow [27].

The usual flow of operations explained before for WFS or WCS will also be useful here. First, the list of available processes can be obtained as part of the service metadata document.

Request: `http://geoprocessing.demo.52north.org/wps/WebProcessingService?service=WPS&request=GetCapabilities`

In this case, Fig. 16.22 shows a fragment of the service metadata that includes an offering of a buffer operation identified as `org.n52.wps.server.algorithm.SimpleBufferAlgorithm`. With this identifier, we can now use the DescribeProcess operation to request more information about the process and, in particular, information on how to formulate a query by including the necessary inputs (Fig. 16.23).

Request: `http://geoprocessing.demo.52north.org/wps/WebProcessingService?service=WPS&request=DescribeProcess&version=2.0.0&identifier=org.n52.wps.server.algorithm.SimpleBufferAlgorithm`

With this response, we learned that the process needs two inputs: *data* and *width* that correspond to the two classical inputs of a buffer operation: the vector data and the buffer distance. Now, we can build a request to execute the buffer operation. Here, we face a problem: the request is too complicated to be encoded in a KVP string. A POST request needs to be invoked. This requires creating an XML document beforehand and sending it to the service in an HTTP POST verb. As *data* for the buffer, we can provide any valid URL to retrieve the data. In this particular example, we will use a WFS service that provides some Tasmanian roads in longitude-latitude and we shall request a buffer of 0.05 degrees. To get the roads, we provide a reference to a GetFeature request to a WFS service: `http://geoprocessing.demo.52north.org:8080/geoserver/wfs?SERVICE=WFS&VERSION=1.0.0&REQUEST=GetFeature&TYPENAME`

Fig. 16.23 WPS DescribeProcess document response (fragment)

```
<?xml version="1.0" encoding="UTF-8"?>
<wps:ProcessOfferings >
  <wps:ProcessOffering processVersion="1.1.0" jobControlOptions="sync-execute
async-execute" outputTransmission="value reference">
    <wps:Process>
      <ows:Title>org.n52.wps.server.algorithm.SimpleBufferAlgorithm</ows:Title>
      <ows:Identifier>org.n52.wps.server.algorithm.SimpleBufferAlgorithm
      </ows:Identifier>
      <wps:Input minOccurs="1" maxOccurs="1">
        <ows:Title>data</ows:Title>
        <ows:Identifier>data</ows:Identifier>
        <ns:ComplexData xmlns:ns="http://www.opengis.net/wps/2.0">
          <ns:Format default="true" mimeType="application/vnd.google-
earth.kml+xml" schema="http://schemas.opengis.net/kml/2.2.0/ogckml22.xsd"/>
          <ns:Format mimeType="application/vnd.geo+json"/>
          <ns:Format mimeType="text/xml"
schema="http://schemas.opengis.net/gml/2.0.0/feature.xsd"/>
          <ns:Format mimeType="text/xml"
schema="http://schemas.opengis.net/gml/2.1.2/feature.xsd"/>
          <ns:Format mimeType="text/xml; subtype=gml/2.0.0"
schema="http://schemas.opengis.net/gml/2.0.0/feature.xsd"/>
          <ns:Format mimeType="text/xml; subtype=gml/2.1.2"
schema="http://schemas.opengis.net/gml/2.1.2/feature.xsd"/>
        </ns:ComplexData>
      </wps:Input>
      <wps:Input minOccurs="1" maxOccurs="1">
        <ows:Title>width</ows:Title>
        <ows:Identifier>width</ows:Identifier>
        <ns:LiteralData xmlns:ns="http://www.opengis.net/wps/2.0">
          <ns:Format default="true" mimeType="text/plain"/>
          <ns:Format mimeType="text/xml"/>
          <LiteralDataDomain default="true">
            <ows:AnyValue/>
            <ows:DataType ows:reference="xs:double"/>
          </LiteralDataDomain>
        </ns:LiteralData>
      </wps:Input>
      <wps:Output>
        <ows:Title>result</ows:Title>
        <ows:Identifier>result</ows:Identifier>
        <ns:ComplexData xmlns:ns="http://www.opengis.net/wps/2.0">
          <ns:Format default="true" mimeType="application/vnd.google-
earth.kml+xml" schema="http://schemas.opengis.net/kml/2.2.0/ogckml22.xsd"/>
          <ns:Format mimeType="text/xml"/>
          <ns:Format mimeType="text/xml"
schema="http://schemas.opengis.net/gml/2.0.0/feature.xsd"/>
          <ns:Format mimeType="text/xml"
schema="http://schemas.opengis.net/gml/2.1.2/feature.xsd"/>
          <ns:Format mimeType="text/xml; subtype=gml/2.0.0"
schema="http://schemas.opengis.net/gml/3.0.0/base/feature.xsd"/>
          <ns:Format mimeType="text/xml"
          <ns:Format mimeType="application/vnd.geo+json"/>
        </ns:ComplexData>
      </wps:Output>
    </wps:Process>
  </wps:ProcessOffering>
</wps:ProcessOfferings>
```

`=topp:tasmania_roads&SRS=EPSG:4326`
`&OUTPUTFORMAT=GML3`

The XML document shown in Fig. 16.24 is sent to the WPS service `http://geoprocessing.demo.52north.org/wps/WebProcessingService` in an HTTP POST message, which returns a response with the resulting buffer (polygons around the linear roads) in KML format.

Sometimes, executions are too long for the user to wait for the result. WPS offers the possibility to execute asynchronous requests. In this case, the server responds immediately on receiving the request, with instructions on how to get the result later. The client can ask for the status of the process from time to time, until the result is available. After completion, the final result can be retrieved (Fig. 16.25).

Fig. 16.24 WPS Execute XML document

```
<?xml version="1.0" encoding="UTF-8"?>
<wps:Execute
  xmlns:wps="http://www.opengis.net/wps/2.0"
  xmlns:ows="http://www.opengis.net/ows/2.0"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.opengis.net/wps/2.0
http://schemas.opengis.net/wps/2.0/wps.xsd"
  service="WPS" version="2.0.0" response="document" mode="sync">
  <ows:Identifier>org.n52.wps.server.algorithm.SimpleBufferAlgorithm</ows:I
dentifier>
  <wps:Input id="data">
    <wps:Reference
xlink:href="http://geoprocessing.demo.52north.org:8080/geoserver/wfs?SERVIC
E=WFS&amp;VERSION=1.0.0&amp;REQUEST=GetFeature&amp;TYPENAME=topp:tasmania_r
oads&amp;SRS=EPSG:4326&amp;OUTPUTFORMAT=GML3"
schema="http://schemas.opengis.net/gml/3.1.1/base/feature.xsd" />
  </wps:Input>
  <wps:Input id="width">
    <wps:Data>
      <wps:LiteralValue>0.05</wps:LiteralValue>
    </wps:Data>
  </wps:Input>
  <!-- Uses default output format -->
  <wps:Output id="result" transmission="value"/>
</wps:Execute>
```

16.3.6 Discovery Services

Web search engines are not so good at finding geospatial information on the web, so specific catalogues for geospatial data discovery are needed. These catalogues cannot be based on the data itself, which is, in most cases, not self-describing, but on descriptions about the data (called metadata). A discovery service is a metadata catalogue that has a query interface to request the kind of data that the user needs. The catalogue compares the query parameters with the metadata previously stored and creates an answer in the form of a list of items that can potentially be of the interest of the user. OGC has produced several standards for geospatial catalog services: One of the most popular ones is the catalogue service for the web (CSW) [28] based on ISO 19115 metadata [29] (Figs. 16.26–16.29). Some people criticize the ISO metadata, arguing that it is too complex a model, and simpler solutions based on Dublin Core have also been produced. Query languages have also been simplified, and OGC has also proposed an extension of the common OpenSearch query language making it spatiotemporally aware: OpenSearchGEO.

This standard works in a very similar way to WFS, but the resources it deals with are database "records" instead of geospatial "features". Since there are several variants of CSW catalogue standards, the service metadata document provides an overview of the capabilities of this service including the metadata encodings delivered and the filter languages supported.

Request: `http://www.ign.es/csw-inspire/srv/spa/csw?service=CSW&request=GetCapabilities`

In the same way, WFS has a DescribeFeature operation that returns the XML schema (GML application schema) for the requested features; the DescribeRecord operation returns the XML schema for the metadata records. Figure 16.27 shows a service supporting both Dublin Core and ISO 19115 metadata records.

Finally, it is possible to retrieve metadata records in ISO XML encoding by specifying the desired schema (resultType parameter) and providing a filter. The simplest filter that emulates the Google search text query is "Like". In the following KVP request, the CQL query language is selected and the "like" query is formulated with the word "topo" in between % signs to indicate that the text is only a fragment of a work; obviously, we are aiming to discover "TOPOgraphic" maps.

```
http://www.ign.es/csw-inspire/srv/spa/
csw?version=2.0.2&service=CSW&request=
GetRecords&resultType=results
&outputSchema=http://www.isotc211.org/
2005/gmd&constraintlanguage=CQL_TEXT
&constraint=%22csw:AnyText%20Like
%20'%25topo%25'%22
```

The result is a collection of records that, in this case, is retrieved as a ISO 19139 XML document (Fig. 16.28).

Catalogues have some other characteristics, such as the possibility to harvest other catalogues, which are not explained here.

16

Fig. 16.25 WPS Execute document response (fragment)

```xml
<?xml version="1.0" encoding="UTF-8"?>
<wps:Result>
  <wps:JobID>7262fb43-d48d-4f23-afe3-5bce0cc7811b</wps:JobID>
  <wps:Output id="result">
    <wps:Data schema="http://schemas.opengis.net/kml/2.2.0/ogckml22.xsd"
mimeType="application/vnd.google-earth.kml+xml">
      <kml:kml xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:kml="http://earth.google.com/kml/2.1">
        <kml:Document id="featureCollection">
          <kml:Placemark id="ID1">
            <kml:MultiGeometry>
              <kml:Polygon>
                <kml:outerBoundaryIs>
                  <kml:LinearRing>
                    <kml:coordinates>146.57855346804192,-41.201323558388665
146.64342790226317,-41.20524509998472 146.65545052760174,-
41.207469488788114 146.66657384077183,-41.21254524781518
146.7890489128575,-41.28774800887231 146.81360198036379,-41.29809249796923
146.8220893899868,-41.30267821480601 146.8459936509968,-41.318752196713696
146.8729421418907,-41.33001622177321 146.89795867662596,-41.32947411713267
146.92159004450176,-41.32850862249745 146.99446995649507,-
41.307955962084804 147.00306754930838,-41.30632696699412
147.0118174734002,-41.30622182149322 147.10211947340022,-41.313061821493214
147.11177356253248,-41.31475656660984 147.12091152267854,-
41.318302167031845 147.12918218715467,-41.32356236732341
147.13626771896324,-41.330335020937284 147.14189582509047,-
41.33835985859323 147.1458502205742,-41.34732849028052 147.14797894021362,-
41.35689625651306 147.14820017850678,-41.36669547340021
147.14650543339016,-41.37634956253247 147.14295983296816,-
41.385487522678545 147.13769963267657,-41.39375818715466
147.1309269790627,-41.40084371896324 147.12290214140677,-41.406471825090485
147.1139335097195,-41.41042622057421 147.10436574348694,-41.412554940213624
147.09456652659978,-41.41277617850678 147.0130918009065,-41.40660480773159
146.94307504350493,-41.4263500379152 146.931545121404,-41.42818532086264
146.901562121404,-41.42941032086264 146.90060423836542,-41.42944026456924
146.86451723836544,-41.43022226456924 146.85413235091937,-41.42936117500916
146.84415148008952,-41.4263662493046 146.8028914800895,-41.4091202493046
146.7942736100132,-41.40447978519399 146.77031008589927,-41.38836595289898
146.7467160196362,-41.37842550203077 146.73996615922817,-41.37495675218482
146.62490795802924,-41.30430813557185 146.57175109773684,-41.30109490001527
146.57021576113064,-41.300978339985946 146.46402976113063,-
41.29127033998595 146.45440322748223,-41.289425496388134
146.44532157611016,-41.28573805739338 146.43713380978357,-
41.280349729214336 146.4301545797707,-41.27346758228109 146.424652093966 8,-
41.265356093640754 146.42083780981565,-41.2563269832528
146.41885830811802,-41.246727234765835 146.41878966001406,-
41.23692576113063 146.42063450361186,-41.22729922748224 146.4243219426066,-
41.21821757611017 146.42971027078568,-41.21002980978359 146.4365924177189,-
41.203050579770675 146.44470390635925,-41.19754809396683
146.4537330167472,-41.19373380981566 146.46333276523416,-41.19175430811803
146.47313423886936,-41.19168566001405 146.57855346804192,-
41.201323558388665</kml:coordinates>
                  </kml:LinearRing>
                </kml:outerBoundaryIs>
              </kml:Polygon>
            </kml:MultiGeometry>
          </kml:Placemark>
        </kml:Document>
      </kml:kml>
    </wps:Data>
  </wps:Output>
</wps:Result>
```

Fig. 16.26 CSW service metadata document response (fragment)

```xml
<?xml version="1.0" encoding="UTF-8"?>
<csw:Capabilities>
  <ows:ServiceIdentification>...</ows:ServiceIdentification>
  <ows:ServiceProvider>...</ows:ServiceProvider>
  <ows:OperationsMetadata>
    ...
    <ows:Operation name="GetRecords">
      <ows:Parameter name="outputSchema">
        <ows:Value>http://www.opengis.net/cat/csw/2.0.2</ows:Value>
        <ows:Value>http://www.isotc211.org/2005/gmd</ows:Value>
      </ows:Parameter>
      <ows:Parameter name="typeNames">
        <ows:Value>csw:Record</ows:Value>
        <ows:Value>gmd:MD_Metadata</ows:Value>
      </ows:Parameter>
      <ows:Parameter name="CONSTRAINTLANGUAGE">
        <ows:Value>FILTER</ows:Value>
        <ows:Value>CQL_TEXT</ows:Value>
      </ows:Parameter>
      <ows:Constraint name="SupportedISOQueryables">
        <ows:Value>Operation</ows:Value>
        <ows:Value>Format</ows:Value>
        <ows:Value>OrganisationName</ows:Value>
        ...
        <ows:Value>Abstract</ows:Value>
      </ows:Constraint>
      <ows:Constraint name="AdditionalQueryables">
        <ows:Value>Relation</ows:Value>
        <ows:Value>AccessConstraints</ows:Value>
        <ows:Value>ResponsiblePartyRole</ows:Value>
        <ows:Value>OnlineResourceMimeType</ows:Value>
        <ows:Value>OnlineResourceType</ows:Value>
        ...
      </ows:Constraint>
    </ows:Operation>
  </ows:OperationsMetadata>
  <ogc:Filter_Capabilities>
    <ogc:Spatial_Capabilities>
      <ogc:GeometryOperands>
        <ogc:GeometryOperand>gml:Envelope</ogc:GeometryOperand>
        <ogc:GeometryOperand>gml:Point</ogc:GeometryOperand>
        <ogc:GeometryOperand>gml:LineString</ogc:GeometryOperand>
        <ogc:GeometryOperand>gml:Polygon</ogc:GeometryOperand>
      </ogc:GeometryOperands>
      <ogc:SpatialOperators>
        <ogc:SpatialOperator name="BBOX" />
        <ogc:SpatialOperator name="Equals" />
        <ogc:SpatialOperator name="Overlaps" />
        <ogc:SpatialOperator name="Disjoint" />
        <ogc:SpatialOperator name="Intersects" />
        <ogc:SpatialOperator name="Touches" />
        <ogc:SpatialOperator name="Crosses" />
        <ogc:SpatialOperator name="Within" />
        <ogc:SpatialOperator name="Contains" />
      </ogc:SpatialOperators>
    </ogc:Spatial_Capabilities>
    <ogc:Scalar_Capabilities>
      <ogc:LogicalOperators />
      <ogc:ComparisonOperators>
        <ogc:ComparisonOperator>EqualTo</ogc:ComparisonOperator>
        <ogc:ComparisonOperator>Like</ogc:ComparisonOperator>
        <ogc:ComparisonOperator>LessThan</ogc:ComparisonOperator>
        <ogc:ComparisonOperator>GreaterThan</ogc:ComparisonOperator>
        <ogc:ComparisonOperator>LessThanEqualTo</ogc:ComparisonOperator>
        <ogc:ComparisonOperator>GreaterThanEqualTo</ogc:ComparisonOperator>
        <ogc:ComparisonOperator>NotEqualTo</ogc:ComparisonOperator>
        <ogc:ComparisonOperator>Between</ogc:ComparisonOperator>
      </ogc:ComparisonOperators>
    </ogc:Scalar_Capabilities>
    <ogc:Id_Capabilities>
      <ogc:EID />
      <ogc:FID />
    </ogc:Id_Capabilities>
  </ogc:Filter_Capabilities>
</csw:Capabilities>
```

Fig. 16.27 CSW service metadata document response (simplified)

```
<?xml version="1.0" encoding="UTF-8"?>
<csw:DescribeRecordResponse>
  <csw:SchemaComponent targetNamespace="http://www.opengis.net/cat/csw/2.0.2"
schemaLanguage="http://www.w3.org/XML/Schema">
    <xsd:schema version="2.0.2">
      <xsd:import namespace="http://purl.org/dc/terms/" schemaLocation="rec-
dcterms.xsd" />
      <xsd:import namespace="http://purl.org/dc/elements/1.1/"
schemaLocation="rec-dcmes.xsd" />
      <xsd:complexType name="SummaryRecordType" final="#all">
        <xsd:complexContent>
          <xsd:extension base="csw:AbstractRecordType">
            <xsd:sequence>
              <xsd:element ref="dc:identifier" minOccurs="1" />
              <xsd:element ref="dc:title" minOccurs="1" />
              <xsd:element ref="dc:type" minOccurs="0" />
              <xsd:element ref="dc:subject" minOccurs="0 />
              <xsd:element ref="dc:format" minOccurs="0" />
              <xsd:element ref="dc:relation" minOccurs="0" />
              <xsd:element ref="dct:modified" minOccurs="0" />
              <xsd:element ref="dct:abstract" minOccurs="0" />
              <xsd:element ref="dct:spatial" minOccurs="0" />
              <xsd:element ref="ows:BoundingBox" minOccurs="0" />
            </xsd:sequence>
          </xsd:extension>
        </xsd:complexContent>
      </xsd:complexType>
    </xsd:schema>
  </csw:SchemaComponent>
  <csw:SchemaComponent targetNamespace="http://www.opengis.net/cat/csw/2.0.2"
schemaLanguage="http://www.w3.org/XML/Schema">
    <xs:schema version="0.1">
      <xs:import namespace="http://www.isotc211.org/2005/gco"
schemaLocation="../gco/gco.xsd" />
      <xs:complexType name="AbstractMD_Identification_Type" abstract="true">
        <xs:complexContent>
          <xs:extension base="gco:AbstractObject_Type">
            <xs:sequence>
              <xs:element name="citation" type="gmd:CI_Citation_PropertyType" />
              <xs:element name="abstract" type="gco:CharacterString_PropertyType"
/>
              <xs:element name="purpose" type="gco:CharacterString_PropertyType"
minOccurs="0" />
              <xs:element name="credit" type="gco:CharacterString_PropertyType"
minOccurs="0" maxOccurs="unbounded" />
              <xs:element name="status" type="gmd:MD_ProgressCode_PropertyType"
minOccurs="0" maxOccurs="unbounded" />
              <xs:element name="pointOfContact"
type="gmd:CI_ResponsibleParty_PropertyType" minOccurs="0" maxOccurs="unbounded" />
              <xs:element name="resourceMaintenance"
type="gmd:MD_MaintenanceInformation_PropertyType" minOccurs="0"
maxOccurs="unbounded" />
              <xs:element name="graphicOverview"
type="gmd:MD_BrowseGraphic_PropertyType" minOccurs="0" maxOccurs="unbounded" />
              <xs:element name="resourceFormat" type="gmd:MD_Format_PropertyType"
minOccurs="0" maxOccurs="unbounded" />
              <xs:element name="descriptiveKeywords"
type="gmd:MD_Keywords_PropertyType" minOccurs="0" maxOccurs="unbounded" />
              <xs:element name="resourceSpecificUsage"
type="gmd:MD_Usage_PropertyType" minOccurs="0" maxOccurs="unbounded" />
              <xs:element name="resourceConstraints"
type="gmd:MD_Constraints_PropertyType" minOccurs="0" maxOccurs="unbounded" />
              <xs:element name="aggregationInfo"
type="gmd:MD_AggregateInformation_PropertyType" minOccurs="0"
maxOccurs="unbounded" />
            </xs:sequence>
          </xs:extension>
        </xs:complexContent>
      </xs:complexType>
    </xs:schema>
  </csw:SchemaComponent>
</csw:DescribeRecordResponse>
```

Fig. 16.28 CSW service metadata document response (fragment)

```xml
<?xml version="1.0" encoding="UTF-8"?>
<csw:GetRecordsResponse>
  <csw:SearchStatus timestamp="2018-08-26T07:36:21" />
  <csw:SearchResults numberOfRecordsMatched="5110" numberOfRecordsReturned="10"
elementSet="summary" nextRecord="11">
    <gmd:MD_Metadata>
      <gmd:fileIdentifier>
        <gco:CharacterString>spaignCartografiaHistoricaIGN201308020055
        </gco:CharacterString>
      </gmd:fileIdentifier>
      ...
    </gmd:MD_Metadata>
    ...
  </csw:SearchResults>
</csw:GetRecordsResponse>
```

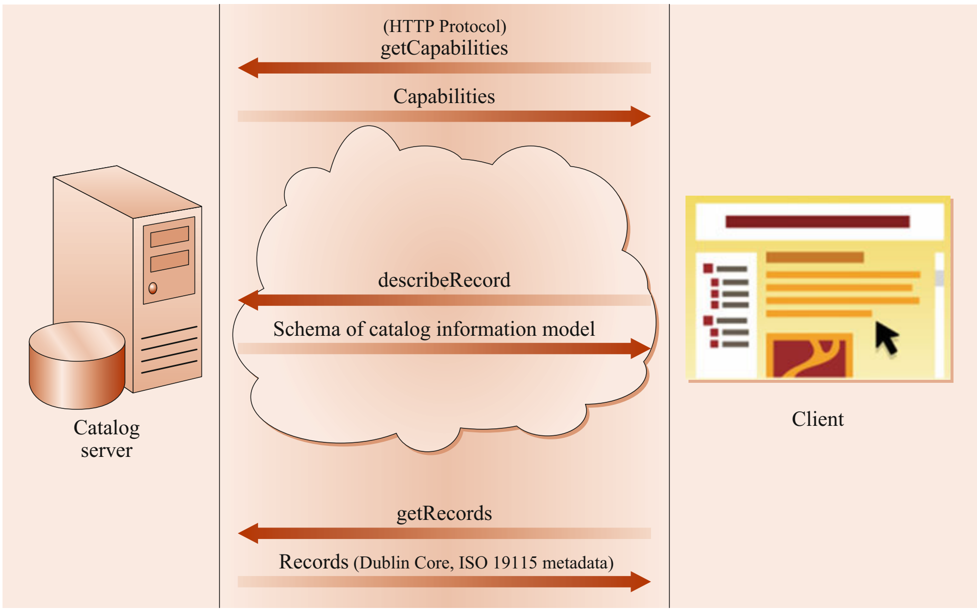


Fig. 16.29 Catalogue service for web operations

16.4 Integrated Geospatial Client

The geospatial web services described in the previous sections are pieces of the puzzle that need to be put together when developing a geospatial client interface with a user. Users of distributed geospatial information should not know about the complexities of the application protocol, and they should have the illusion that all happens in a single environment. They should not normally realize that they are accessing distributed data except, perhaps, if there are internet access problems. To achieve this, an integrated client should automate the service connections and should be able to handle different protocols and formats in a transparent way.

Figure 16.30 presents a diagram of an integrated client that uses WMS, WFS, and SOS requests simultaneously. WMS is used to get orthophotographic imagery, WFS provides nautical charts, and SOS is used to get positions of some whale observations. WMS GetMap, WFS GetFeature, and SOS GetFeatureOfInterest are formulated for their respective services, each one returning a different format. The client is able to interpret the document formats and present the geospatial information overlapped on the screen. Each time that the user does a zoom or a pan, the whole cycle is repeated. The client should be designed as a multithread application where WMS, WFS, and SOS requests can happen simultaneously, and they are presented on the screen as soon as one of the responses arrives. Due to different delays in the network, sometimes, the SOS observations are represented first, and other times the WMS imagery appears first, but in the end, all the necessary information will emerge, and the user will have a more fluid experience with the application. If the integrated client is a JavaScript map browser, implementing multithread applications is relatively easy to program by combining the following strategies: the `<img>` elements are loaded in an independent thread. `GetTimeOut()` functions are executed in a new thread and AJAX requests can be asynchronous.

16

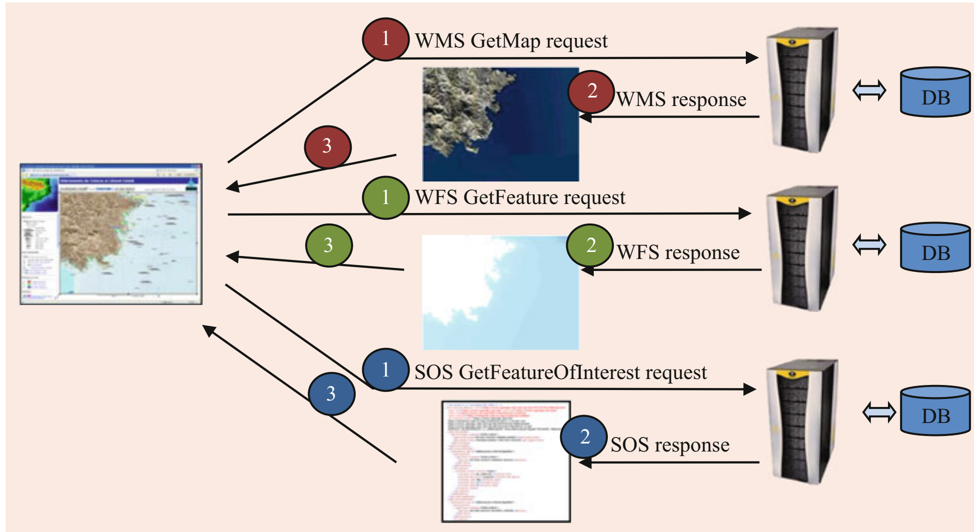


Fig. 16.30 Integrated client with WMS, WFS, and SOS

Today, modern GIS applications can also act as geospatial integrated clients, and data coming from distributed geospatial services can be combined with local data in an integrated way.

OGC recently released a standard created to allow a set of configured information resources (service set) to be passed between applications, primarily as a collection of services as well as in-line content. The goal of OGC OWS context is to support the distribution of search results or the exchange of a set of visualized resources, such as WFS, WMS, WMTS, WCS, and other services included in a *common operating picture*. Additionally, OWS Context can deliver a set of configured processing services (WPS) and the parameters to allow the reproduction of the processing on different nodes. OWS Context replaces a previous OGC attempt at providing such a capability (the web map context, WMC), which was limited to WMS.

16.5 Deploying Web Services

If you have data to offer, you can deploy your own geospatial server. There are many vendors that provide software solutions that implement web services. As a starting point we need a web server where we can deploy the geospatial web server application. The operating system of modern desktop computers can incorporate a web server. In Linux, Apache is the usual alternative and the Internet Information Server is by default a part of modern versions of Windows. In our experience, adding a geospatial web service to them requires some knowledge on how a web server works and several configuration steps. Instead of using the current web server already installed, it is easy to look for a download package that incorporates their own web server already integrated and configured with the geospatial web server. To illustrate how to set up a service, we will user GeoServer, which is an open-source implementation for WMS, WMTS, WCS, and WFS servers, also having extensions for CSW and WPS. GeoServer offers a Windows distribution with a integrated web server that is very easy to set up and configure. At the time of writing these lines, the GeoServer official URL for download is http://geoserver.org/release/stable/. The only prerequisite in GeoServer is that it needs the Java Development Kit installed on the machine before being able to run it (JDK is not the same as the JRE that is commonly used by many other applications).

Actually, GeoServer default setup automatically starts a geospatial server that includes some sample data. This is useful to demonstrate the correct performance of your new installation but does not help in publishing your own data. GeoServer offers a web interface to configure the data that you want to display as a service. To publish a shape file as a WMS service you should take the following steps (Fig. 16.31)

1. Generate a new *workspace* for your data. This will separate your "things" from the demo data that is already included.
2. Copy your shape file in the `(geoserver install)/data_dir/data/`
3. Create a new *data source*. GeoServer supports many data sources, both files and geodatabases. A shapefile is one of the options under *vector data sources*.
4. Create a new *style*. This is done by providing an SLD file that indicates how data should be rendered.
5. Create a new *layer*. In this process, you select a data source and associate it to a style.
6. Test your new *layer* as a web service in the *layer preview* option

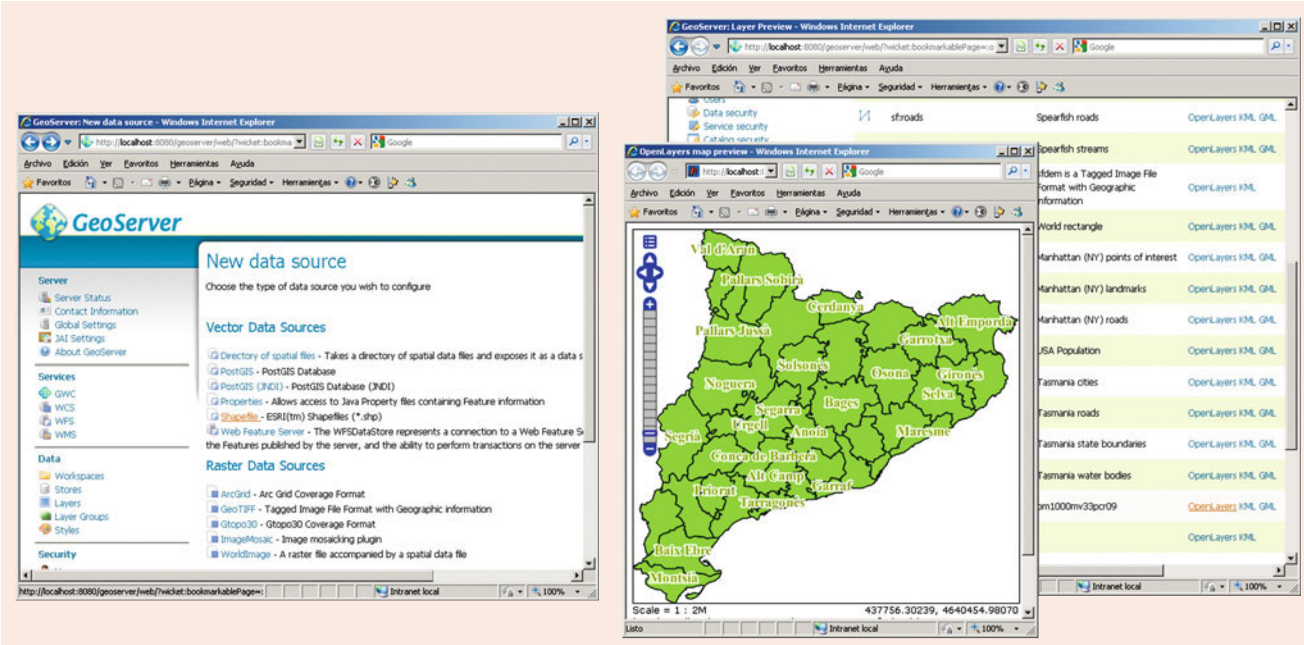


Fig. 16.31 Two steps to configure a layer in a WMS in GeoServer. Setting a data source and previewing the result (the administrative limits of Catalonia were produced by the Institut Cartogràfic i Geològic de Catalunya)

Now, your geospatial web server has been set and is ready to be used by a standards client. The URL of your WMS service is `http://my.server.name:8080/geoserver/wms` and the data is available as other services, such as a WFS service in `http://my.server.name:8080/geoserver/wfs`.

16.6 The Emergence of APIs

The services presented in the previous sections are part of the OGC Web Services approach that define implementation level communication protocols between client and server on top of the HTTP application layer. For web developers, it is more natural to import an application program interface (API) than to understand a new communication layer. Originally, the term API was used as a synonym of "a library to create user interfaces". Today, the term is still used in this sense also for libraries that facilitate the creation of map browser user interfaces. However, the same term has recently been applied to well-documented RESTful interfaces that present lists of resources and operations. These two different approaches are explained in the following two Sects. 16.6.1 and 16.6.2.

16.6.1 Geospatial APIs for Creating User Interfaces

Integrated clients implement the complexities of geospatial services and protocols in such a way that final users do not need to be aware of the complexities of the distributed computing architecture. In the same way, some software libraries provide APIs warping the complexities of geospatial services and protocols for software developers too. In this case, the software developers use an API to create map applications that run on web browsers. This type of API provides a set of JavaScript functions to simplify the creation of a user interface for presenting and querying geospatial data in a web page. Internally, the library uses geospatial web communications (open standards or proprietary protocols), but developers only need to know the URL of the services and the layers available. By using the API, they will not need to implement the service protocol themselves, as the API will execute it for them.

To illustrate how this kind of user interface APIs work, we present a small example with Leaflet (https://leafletjs.com/reference-1.7.1.html). Leaflet is a JavaScript API for building interactive map browsers that is able to communicate with many services, including OGC WMS services. In Fig. 16.32, we can see a fragment of code where the API is imported by adding a style-sheet (leaflet.css) and a script (leaflet.js). The API interacts with a `<div>` area defined in an HTML page and draws a map inside. The first line of code creates an empty "map" in a region of the Earth defined by some coordinates situated over Barcelona and associates it to the `<div>` identifier. The second line of code adds a WMS layer to the "map" using the URL of the server (from the Catalan Cartographic and Geological Institute), the layer name, and the file format as input. As of July 2021, the current Leaflet version was 1.7.1, but the example is backwards

```
<html>
<head>
        <link rel="stylesheet" href="https://unpkg.com/leaflet@1.7.1/dist/leaflet.css" />
</head>
<body>
        <div id="map" style="width: 600px; height: 400px"></div>
        <script src="https://unpkg.com/leaflet@1.7.1/dist/leaflet.js"></script>
        <script>
                var map = L.map('map').setView([41.90, 2.15], 7);
        L.tileLayer.wms("http://geoserveis.icc.cat/icc_mapesbase/wms/service", {
                    layers: 'orto5m',
                    format: 'image/png',
                    transparent: true,
                    attribution: "ICGC © 2021"
                }).addTo(map);
        </script>
</body>
</html>
```

Fig. 16.32 Leaflet API used in a JavaScript code that internally uses the WMS protocol (fragment)

compatible, back to version 0.7.3 or lower. The library manages the actions of the user on the map and transforms them into client–server communication messages. Web developers do not need to be aware of the communication details and only need to know how JavaScript works and how to use the library.

16.6.2 Geospatial Web APIs

This kind of API provides a completely different approach. In this case, the implementation level protocol is completely removed, and the HTTP application level protocol in an ROA style (RESTful) is used instead. The focus is on defining resources (associating them to URIs) that are transmitted from or to a server using HTTP verbs (with the meaning specified in Sect. 16.2.4). In the end, the API defines a list of URI templates for resources, the valid values for each variable in the URI templates, and which HTTP verbs can be applied to each resource. All this information can be formally presented in an API description document. The OpenAPI 3.0 document format (formerly known as Swagger) is often used as a standard way to provide API descriptions.

The OGC has embraced a process of transforming the OGC Web Service standards (WMS, WFS, WCS, WPS, etc.) into an OGC Web API family of standards that takes advantage of modern web development practices. These standards are built in a coordinated way that will allow for a single integrated API implementation to potentially combine characteristics of all the OGC Web API standards. At the same time, the OGC Web API standards are constructed as a set of small "building blocks" that can be combined with other functionalities and resources (including others not specified by the OGC) to assemble novel APIs for accessing a combination of geospatial and nongeospatial content. These building blocks are defined not only by the requirements of the specific standards but also through interoperability prototyping and testing in OGC's Innovation Program.

As of July 2021, the OGC API – Features (the evolution of the WFS) has been approved by the OGC, but many other groups are working in the same direction on several API standard candidates such as the OGC API – Tiles (the evolution of the WMTS), the OGC API – Maps (the evolution of the WMS), the OGC API – Records (the evolution of the CSW), the OGC API – Coverages (the evolution of the WCS), the OGC API – Styles (the evolution of the SLD), the OGC API – Processes (the evolution of the WPS), and the Environmental Data Retrieval (a new convenience API focusing on retrieving environmental data inside points, areas, cubes, trajectories, or corridors). The up-to-date panorama of the OGC Web APIs can be found at the initiative page: https://ogcapi.ogc.org.

The OGC Web API family uses a modular approach where each building block is defined as a set of requirements being part of a requirements class. For convenience, some requirements classes are grouped in documents that are called "Parts". Each OGC Web API standard is composed by one or more parts. For example at the time of writing these lines the OGC API – Features has released 2 parts: OGC API – Features – Part 1: Core (OGC 17-069r3) [31] and OGC API – Features – Part 2: Coordinate Reference Systems by Reference (OGC 18-058) [32], and more are expected to be approved soon. To ensure that all building blocks of the OGC Web API are compatible and efforts are well aligned, the OGC is standardizing the OGC API – Common that provides fundamental building blocks for many APIs and have developed OGC Web API Guidelines [30] to follow the OGC Web API standards.

In the OGC Web APIs, there are three main ways of discovering resources: an API description document, the "follow the links" approach, and the predictable common paths. An OpenAPI 3.0 document (https://swagger.io/specification/)

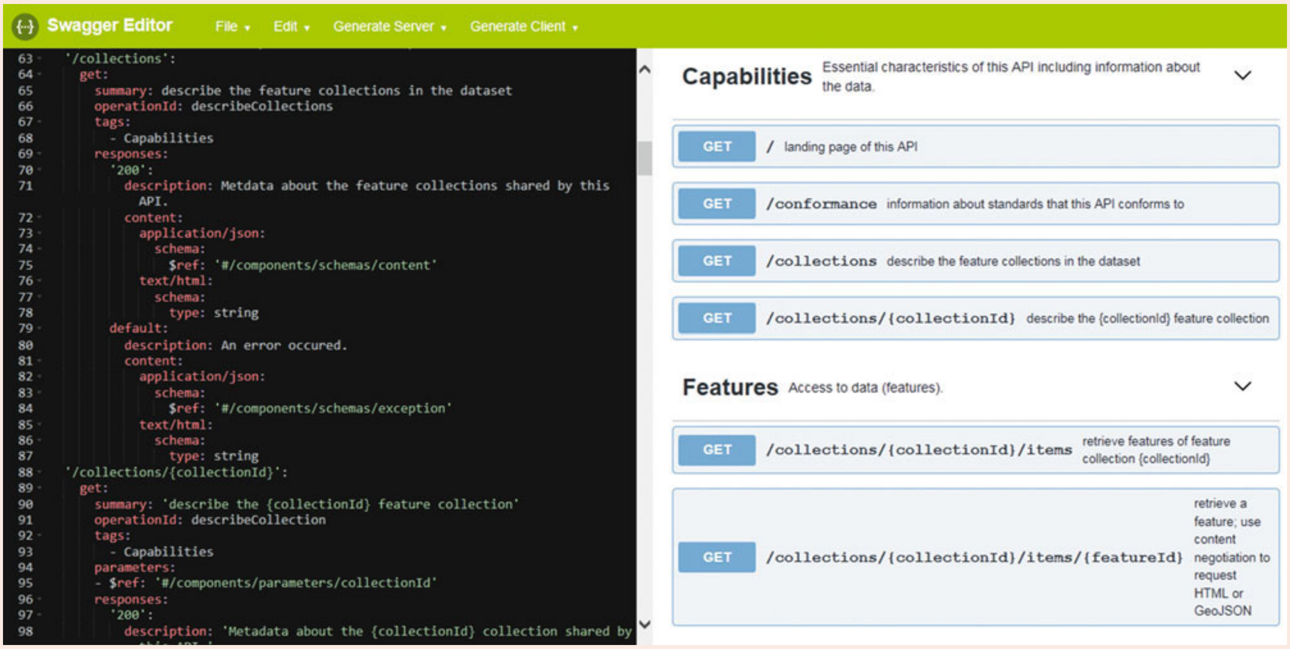


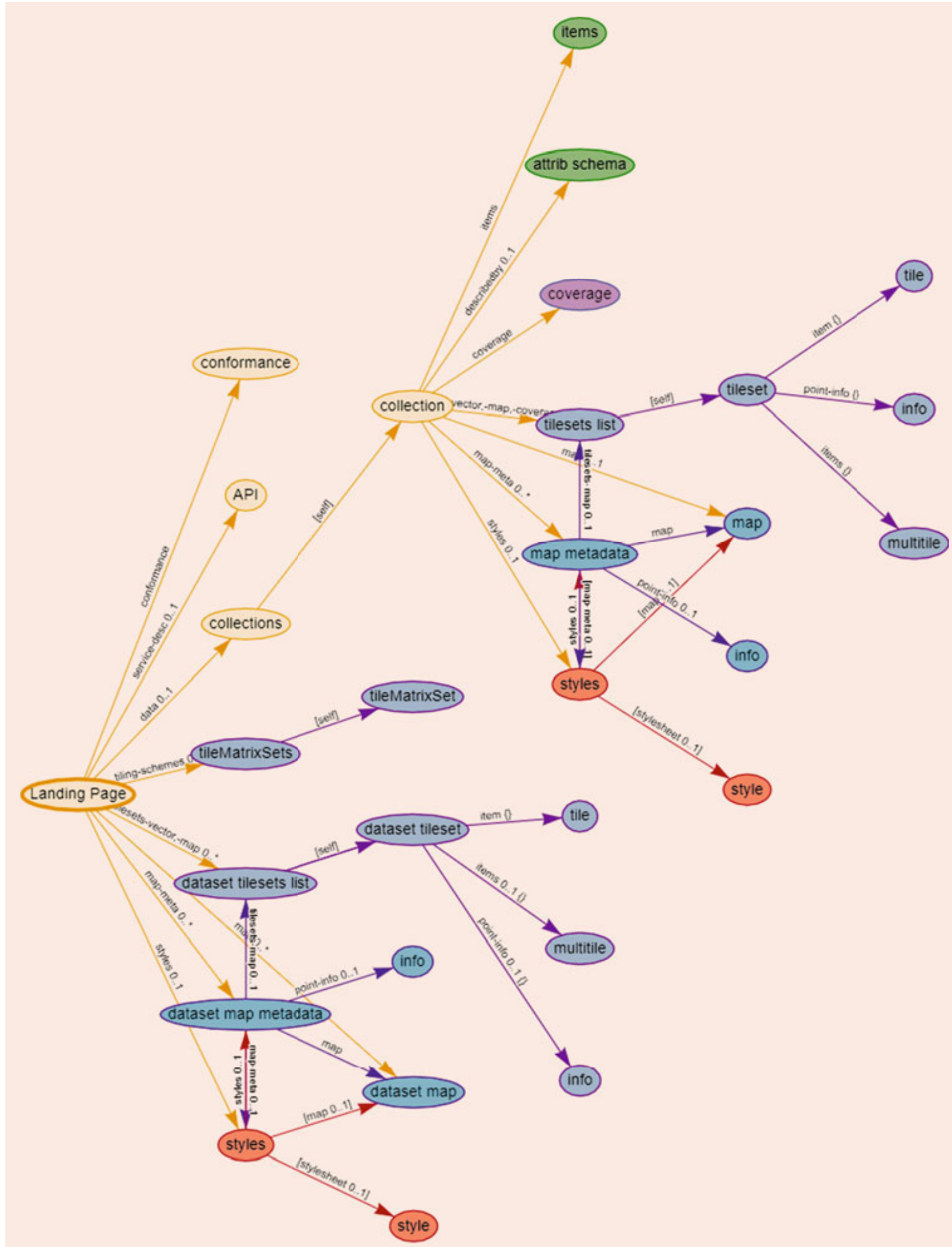
Fig. 16.33 OpenAPI 3.0 document example for an Web API implementing OGC API – Features

Fig. 16.34 A fragment of a JSON representation of a landing page conformal to the OGC API – Features

```
{
    "title": "Buildings in Bonn",
    "description": "Access to data about buildings in the city of Bonn via a Web
API that conforms to the OGC API Features specification.",
    "links": [
        {
            "href": "http://data.example.org/api",
            "rel": "service-desc",
            "type": "application/vnd.oai.openapi+json;version=3.0",
            "title": "the API definition"
        },
        {
            "href": "http://data.example.org/conformance",
            "rel": "conformance",
            "type": "application/json",
            "title": "OGC API conformance classes implemented by this server"
        },
        {
            "href": "http://data.example.org/collections",
            "rel": "data",
            "type": "application/json",
            "title": "Information about the feature collections"
        }
    ]
}
```

is a useful API description document (encoded in JSON or YAML) that lists the available resources in the form of URI templates associated to the HTTP verbs supported for each kind of resource (Fig. 16.33). OpenAPI 3.0 is the only specified API description document included so far in the current OGC Web API standards. The "follow the links" approach uses links included in many resources starting at the landing page, in the same way that HTML web pages do. Figure 16.34 shows a fragment of the landing page encoded in JSON that contains some links including one pointing to the list of "collections". The "collections" resource will include more links and the linked resources even more links, and so on. In particular, the "collections" resource lists the identifiers of each and every available "collection" and provides links pointing to all of them. For each link, there is a "rel" property that describes the type and purpose of the relation.

Fig. 16.35 OGC Web API resources (*ellipses*) and link relations (*arrows*) graph



Possible values for the "rel" property can be found in the IANA link relation types web page (https://www.iana.org/assignments/link-relations/link-relations.xhtml. E.g.: "collection") or in the OGC Naming Authority server (http://www.opengis.net/def/rel. E.g.: "http://www.opengis.net/def/rel/ogc/1.0/conformance"). Figure 16.35 shows a graph representing a set of relations included by several OGC Web API standards, informs on how resources (represented as ellipses) are linked together, and presents the link relation types used for each link in a label next to the arrows connecting the ellipses. The third way to discovering resources is to try out the common paths and URI templates that are exemplified in the different OGC Web API standards. Most of those paths are optional, and server API developers can decide to favor another configuration. However, it is foreseen that most developers will use these common URI templates as the basis for their implementations. In some cases, such as the OGC API – Features – Part 1, these fixed paths are considered mandatory.

"Collection" is one of the most important resources in the OGC Web API family of standard because this shared resource gives access to a group of geospatial data available thought the API in several forms. Depending on the number of OGC Web APIs implemented by an API instance (that are

Table 16.4 Correspondence between OGC Web API URIs and OWS Web Service operations

OGC Web API		OWS Web Service
URI template	Resource retrieved	Correspondence
/	Landing page	OWS Common GetCapabilities ServiceProvider section
/api	Open API document	OWS Common GetCapabilities OperationsMetadata section
/conformance	List of standard and profiles that this service is honoring	OWS Common GetCapabilities ServiceIndentification section
/collections	Enumeration of "groups" of the data ("collection") present in service	WFS GetCapabilities Content section (FeatureType grouping concept)
/collections/{collectionsId}	Details about a collection	WFS DescribeFeatureType response
/collections/{collectionId}/items	Feature collection description	WFS GetFeature operation selecting a TypeName
/collections/{collectionId}/items/{featureId}	Feature	WFS GetFeatureById response
/collections/{collectionId}/tiles/ /{tileMatrixSetId/{tileMatrix}/{tileRow}/{tileCol}	Tile	WMTS GetTile response
/collections/{collectionId}/styles/{styleId}/map	Map	WMS GetMap response
/collections/{collectionId}/coverage	Coverage	WCS GetCoverage response
/collections/{collectionId}/coverage/domainSet	CIS domainSet of a coverage	WCS DescribeCoverage domainSet section
/collections/{collectionId}/coverage/rangeType	CIS rangeType of a coverage	WCS DescribeCoverage rangeType section
/styles/{styleId}	SLD/SE style description	WMS SLD GetStyles

Table 16.5 The tile path fragment building block combined with other resources

URI template	Meaning	OWS Correspondence
/collections/{collectionId}/ /tiles/{tileMatrixSetId/{tileMatrix}/{tileRow}/{tileCol}	Retrieves a vector tile	WMTS GetTile specifying a vector tile format
collections/{collectionId}/map/{Styles)/ /tiles/{tileMatrixSetId/{tileMatrix}/{tileRow}/{tileCol}	Retrieves a map tile	WMTS GetTile common use
collections/{collectionId}/coverage/ /tiles/{tileMatrixSetId/{tileMatrix}/{tileRow}/{tileCol}	Retrieves a coverage tile	N/A

listed in the conformance page) and the data capabilities, a "collection" can be retrieved as:

- Simple features (e.g., in a vector format)
- A coverage (e.g., in raster format; if it is a regular grid coverage)
- Maps (by subsetting a styled map representation)
- Tiles (by dividing the data in a tileset).

Table 16.4 gives an overview of some of the fixed paths suggested by the OGC Web API family of standards.

Even if many URI templates have a logical correspondence to an equivalent OGC Web Service functionality, there are many practical differences. To cite one difference, OGC Web API favors the use of JSON encodings instead of XML encodings. For example, the OGC API – Features – Part 1 uses GeoJSON as the default format (it still provides optional conformance classes for two XML-based GML simple features profiles 0 and 2). In addition, some OGC Web API standards use the concept of building blocks to specify fragments of resource paths. For example, the path fragment /tiles/{tileMatrixSetId/{tileMatrix}/{tileRow}/{tileCol} can be combined with features to get vector tiles, with a map to get map tiles, or with a coverage to get tiles from a coverage (Table 16.5)

There will be many more functionalities that are not explained in this section, such as the capacity to create resources that are "jobs" that creates new resources by executing "processes" (a capacity provided by the OGC API – Processes.

References

1. Champion, M., Ferris, C., Newcomer, E., Orchard, D. (eds.): Web Services Architecture. W3C, Cambridge (2002). W3C Working Draft 14. November 2002, http://www.w3.org/TR/ws-arch/
2. Peng, Z.R., Tsou, M.H.: Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Network. Wiley, New York, p. 720 (2003)
3. Lucchi, R., Millot, M., Elfers, C.: Resource Oriented Architecture and REST. Assessment of Impact and Advantages on INSPIRE. Ispra: European Communities, Brussels (2008)
4. Döllner, J., Jobst, M., Schmitz, P.: Service Oriented Mapping. Lecture Notes in Geoinformation and Cartography. Springer Nature, Cham (2019)
5. W3C: What is Linked Data. W3C, Cambridge (2010). http://www.w3.org/standards/semanticweb/data

6. Pan, M., Pang, B., Li, C., Wang, D., Min, D., Zhao, D.: Research on web mapping of vector nautical charts based on HTML5. In: Int. Conf. Inf. Sci. Cloud Comput. Companion (ISCC-C), pp. 611–617. IEEE, New York (2013)
7. Deng, M., Di, L.: Facilitating Data-Intensive Research and Education in Earth Science: A Geospatial Web Service Approach. Lambert Academic, Saarbrücken, p. 212 (2010)
8. OGC: OGC Reference Model (ORM). OGC, Wayland (2011). OGC 08-062r7, http://www.opengeospatial.org/standards/orm, http://portal.opengeospatial.org/files/?artifact_id=31112
9. Whiteside, A., Greenwood, J.: OpenGIS Web Service Common Standard 2.0. OGC, Wayland (2010). http://www.opengeospatial.org/standards/is
10. Gregorio, J., Fielding, R., Hadley, M., Nottingham, M., Orchard, D.: RFC 6570, URI Template. Internet Engineering Task Force, Tremont (2012)
11. Stefanakis, E.: Web Mapping and Geospatial Web Services: An Introduction. CreateSpace, North Charleston, p. 168 (2015)
12. de la Beaujardiere, J. (ed.): OGC Web Map Service Interface 1.3.0. OGC, Wayland (2004). OGC Document # 01-109r1, 2004-01-20, available at http://www.opengeospatial.org/standards/is
13. Masó, J., Pomakis, K., Julià, N.: OpenGIS Web Map Tile Service Implementation Standard. OGC, Wayland (2007)
14. Loechel, A.J., Schmid, S.: Comparison of different caching techniques for high-performance web map services. Int. J. Spat. Data Infrastruct. Res. **8**, 43–73 (2013)
15. Sample, T.S., Lup, E.: Tile-Based Geospatial Information Systems – Principles and Practices. Springer, Cham (2010)
16. Lalonde, W. (ed.): Styled Layer Descriptor Implementation Specification 1.0.0. OGC, Wayland (2002). OGC 02-070, available at http://www.opengeospatial.org/standards/is
17. Vretanos, P. (ed.): Web Feature Service Implementation Specification 1.0.0. OGC, Wayland (2003). OGC Document #02-058, available from http://www.opengeospatial.org/standards/is
18. Portele, C.: *OpenGIS® Geography Markup Language (GML) Encoding Standard version 3.2.1*. OGC, Wayland (2007)
19. Davis, S.: GIS for Web Developers. Pragmatic Bookshelf, Dallas, p. 254 (2007)
20. Baumann, P., Hirschorn, E., Masó, J.: OGC Coverage Implementation Schema. OGC, Wayland (2009). OGC 09-146r6
21. Baumann, P.: *OGC® WCS 2.0 Interface Standard–Core*. OGC, Wayland (2009). OGC 09-110r3
22. Zhao, P., Di, L.: Geospatial Web Services: Advances in Information Interoperability. IGI Global, Hershey (2010)
23. Botts, M., Percivall, G., Reed, C., Davidson, J.: OGC® sensor web enablement: Overview and high level architecture. In: GeoSensor Networks, pp. 175–190. Springer, Berlin, Heidelberg (2008)
24. Bröring, A., Stasch, C., Echterhoff, J.: OGC Sensor Observation Service Interface Standard OGC 12-006. OGC, Wayland (2012). http://www.opengis.net/doc/IS/SOS/2.0
25. Connecting Geo Consortium: Best Practices for Using Sensor Web Technology. 52 North Initiative, Münster (2016). https://52north.github.io/sensor-web-tutorial/
26. Alameh, N.: Chaining geographic information web services. IEEE Internet Comput. **7**(5), 22–29 (2003)
27. Mueller, M., Pross, B.: OGC WPS 2.0 Interface Standard OGC 14-065. OGC, Wayland (2014). http://docs.opengeospatial.org/is/14-065/14-065.html
28. Nebert, D., Whiteside, A., Vretanos, P.: *OpenGIS® Catalogue Services Specification Version 2.0.2*. OGC, Wayland (2007). OGC 07-006r1
29. Kresse, W., Fadaie, K.: ISO Standards for Geographic Information. Springer, Berlin, Heidelberg (2004)
30. Matheus, A., Terpstra, F., Masó, J. (ed.): OGC Web API Guidelines (2021), available at https://github.com/opengeospatial/OGC-Web-API-Guidelines
31. Portele, C., Vretanos, P., Heazel, C. (ed.): OGC API – Features – Part 1: Core, 1.0.0. OGC, Wayland (2019). OGC Document #17-069r3, 2019-10-14, available at http://docs.opengeospatial.org/is/17-069r3/17-069r3.pdf
32. Portele, C., Vretanos, P. (ed.): OGC API – Features – Part 2: Coordinate Reference Systems by Reference, 1.0.0. OGC, Wayland (2019). OGC Document #18-058, 2020-11-02, available at http://docs.opengeospatial.org/is/18-058/18-058.pdf

Joan Masó Joan Masó received his PhD from the Universitat Autònoma de Barcelona in 2012. He works at CREAF, where he develops GIS, RS, and web mapping tools. He has contributed to several OGC standards (e.g., OGC 07-057r7 Web Map Tile Services). He currently focuses on web services and web clients for geoinformation for GEOSS and Citizen Science. He received the Gardels award in 2018.

Geosemantic Interoperability and the Geospatial Semantic Web

17

Jean Brodeur

Contents

Abstract

This chapter gives an account of the semantic interoperability of geographic information and the various elements involved in it. After providing some historical background in Sects. 17.1–17.3 introduce some basic notions about how human beings abstract phenomena and represent them in human memory. This is fundamental for understanding the purpose of semantics in interoperability. This leads us to introduce the concept of ontology in Sect. 17.4, where we describe its role with respect to semantics. Based on these notions, semantic interoperability of geographic information (i.e., geosemantic interoperability) is then compared in Sect. 17.5 to a communication process. Typically, the interoperability of geographic information is implemented by way of geospatial data infrastructures (SDI), and Sect. 17.6 describes the implementation of geosemantic interoperability through SDI with the introduction of the geospatial information environment. Standards are fundamental to interoperability for geographic information, and Sect. 17.7 provides an overview of the contribution of standards bodies with respect to geosemantic interoperability, including the development of ontologies. Then, Sect. 17.8 provides a description of the valuable contribution the Semantic Web to semantic interoperability and the description of the geospatial Semantic Web. Finally, the conclusion provides some concluding remarks on the chapter.

Keywords

semantics · interoperability · geographic information · ontologics

17.1 Historical Background

The need to share geospatial data and information has been a requirement since the first geographic information systems. Data were produced by different organizations on different systems and were required by others for their use in different applications. This has been the case with topographic data that underlie spatial analysis for environmental studies, emergencies, inundations, traffic control, security, agriculture, etc.

The last decades have been very productive in terms of the development of the interoperability of geographic information to facilitate sharing of and access to geographic data.

J. Brodeur (✉)
Recherche GéoSémantic
Sherbrooke, Québec, Canada
e-mail: recherchegeosemantic@videotron.ca

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_17

Standards bodies, such as ISO/TC 211 and the Open Geospatial Consortium (OGC) Inc., have established the standard foundation on which it has been developed. Spatial data infrastructures have sprung up in national and international organizations (the Canadian Geospatial Data Infrastructure (CGDI), the National Spatial Data Infrastructure (NSDI), the Infrastructure for Spatial Information in Europe (INSPIRE), the Global Spatial Data Infrastructure (GSDI), etc.).

At the same time, we have witnessed the impressive development and evolution of the Internet and the World Wide Web. The Web began with the very simple publication of Web pages, linked together in some cases. Now, it is composed of portals, services, data, documents, videos, music, etc. It is a source of tremendous social growth. Today, the public contributes to the collection and availability of data, information of various kinds such as wikis [1], Google maps [2], OpenStreetMap [3], etc. Currently, the Web is progressing toward the Semantic Web (or Web 3.0), moving from a Web of documents to a Web of data. The Web can now be compared to a worldwide open database [4] that people can query from their own perspective, understanding, or abstraction of real-world phenomena or events to obtain accurate, detailed, and appropriate answers, and likewise people can communicate with each other.

17.2 What Is Semantics about?

The *Merriam-Webster* dictionary defines "semantics" as "*the study of meaning* and a branch of semiotics dealing with the relations between signs and what they refer to ..." [5]. Semantics is also related to the communication and the meaning of a message conveyed by the signs that compose it. Essentially, the meaning that an individual, for instance, associates to signs establishes the link between the signs and their corresponding phenomenon. This has been widely described by many intellectual people [6–9] using the triad: *referent* (i.e., the phenomenon), *signifier* (i.e., the symbols or the physical stimuli that activate thoughts or concepts), and *signified* (i.e., the thoughts that associate a signifier and referent pair), as represented by the semiotic triangle ([6, 10], Fig. 17.1).

It is important to recognize and acknowledge that there is no direct relationship between the signifier and the referent. This relation is made possible via the thoughts that one has in mind. This is the essence of semantics and the underpinning of good communication and the interoperability of geographic information.

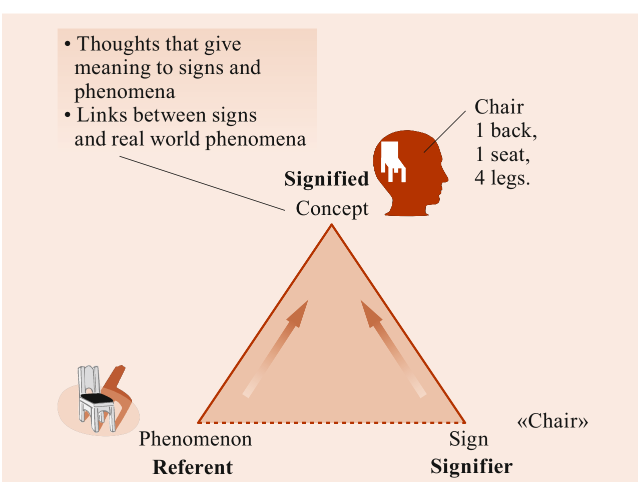


Fig. 17.1 The semiotic triangle

17.3 Semantics Through Cognition

The cognitive development of human beings begins with direct and indirect observation of phenomena. Humans perceive phenomena through their senses, i.e., sight, hearing, smell, touch, and taste. They can also perceive phenomena through technological means such as satellites, digital cameras, microscopes, radar, etc. Perceptions in human beings generate *perceptual states* [11]. A perceptual state is a state of the brain composed of an unconscious neuronal representation originating from the physical input of the human senses. From such perceptual states, human selective attention retains only a significant subset, which is stored as mental images in the long-term memory. A mental image, also called a *perceptual symbol*, constitutes an abstract representation of a phenomenon [11]. There is still an important debate about what form perceptual symbols take in human memory [12, 13]. There are two well-known approaches: the modal and amodal approaches. The modal approach assumes that perceptual symbols are stored in human memory similarly to perceptual states and constitute a representation that is analogous to a picture of the phenomenon (Fig. 17.2). The amodal approach considers that perceptual symbols consist in descriptive forms; i.e., that perceptual states are transformed into propositions describing the set of significant properties of the phenomenon (Fig. 17.3). The objective here is not to add to this debate but to acknowledge the existence of perceptual symbols, which consist of concepts in the human memory that are associated to the phenomenon.

Concepts are arranged in a network in such a manner that similar concepts are attracted together [11, 14]. A concept is used to simulate representations of itself adapted to specific

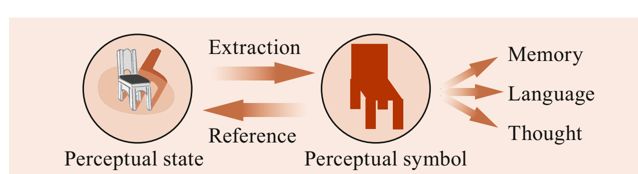


Fig. 17.2 The modal symbol system approach

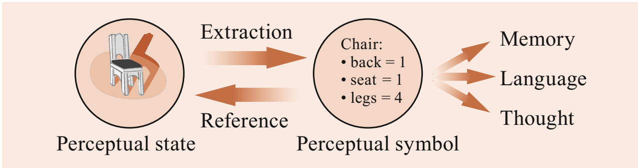


Fig. 17.3 The amodal symbol system approach

contexts and to recognize simulations generated by others' concepts.

As one can recognize, cognition and semantics (as illustrated by the semiotic triangle) are widely interrelated [15]. Referent and signifier are both elements that stimulate human senses to generate ultimately perceptual symbols. Perceptual symbols match concepts (i.e., thoughts) in providing meaning to elements perceived by human senses.

17.4 Ontology

Since Aristotle, philosophers have been concerned with topics such as existence and the knowledge and description of being and truth. These questions have been studied from an ontological perspective. In philosophy, ontology refers to the description of the world [16], and a model and an abstract theory of the world [17]. It is the science of being, the science of *What is?*, the science of the type of entities with their properties, including categories, and relationships that compose reality [16–18]. As there is only one reality to describe, philosophers agree on the existence of one and only one ontology that is independent of any language.

More recently, the term *ontology* has been borrowed by the computer science community, more specifically in artificial intelligence, where *Gruber* [19] defined ontology as *an explicit specification of a conceptualization*. In 1998, *Guarino* went further in defining ontology as [20]:

> a logical theory accounting for the intended meaning of a formal vocabulary – i.e. its ontological commitment to a specific conceptualisation of the world.

In short, an ontology consists in [21]:

> a formal representation of phenomena with an underlying vocabulary including definitions and axioms that make the intended meaning explicit and describe phenomena and their relationships.

In this realm, an ontology consists in the layer that defines concepts of reality [22]. It is an element or a resource from which agents perceive the world [23]. It is made of a specific vocabulary and relations used to describe some aspects of reality, including a set of axioms related to the intended meaning of the vocabulary. In practice, an ontology can take many forms: keyword hierarchy, conceptual schema, semantic network, taxonomy, thesaurus, etc. [24].

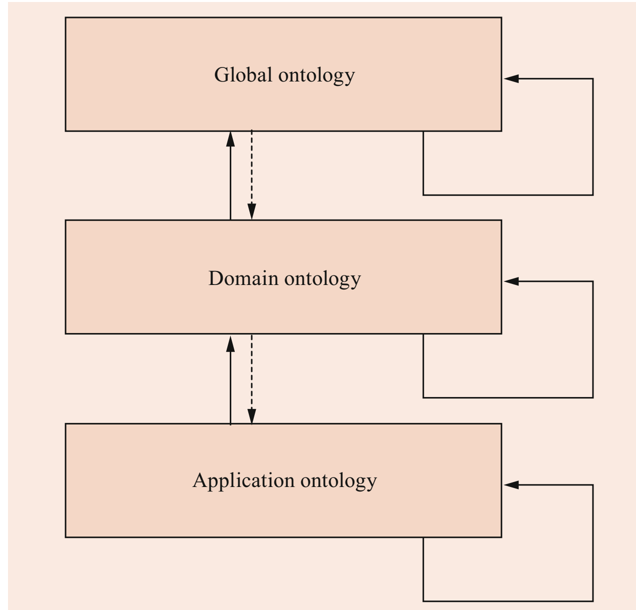


Fig. 17.4 Levels of ontology

The above description and definition of ontology comply with the representation of semantics (Fig. 17.1), since ontology is composed of a set of concepts that associates the symbols or vocabulary with phenomena providing appropriate meaning.

People abstract phenomena of reality from different perspectives and at different levels of detail, depending on how accurate a description is required. Consequently, concepts in ontologies may be depicted from a more general and global purpose to a very specialized context or a specific application. Typically, three levels of ontology are considered: global ontology [20, 22, 25–29], domain ontology [20, 22, 27, 28, 30, 31], and application ontology ([20, 26, 27, 32, 33], Fig. 17.4). Each level is characterized by a specific granularity at which phenomena are abstracted and depicted.

A *global ontology* can be compared to a dictionary, providing definition and meaning of generic terms for general usage (WordNet [34], *Merriam-Webster* dictionary [5], etc.). A lexicon, i.e., a brief dictionary specifically oriented to a given science or technology, is a good example of a *domain ontology*. The *IEEE Standard Computer Dictionary* [35] is a good example of a lexicon. Finally, a glossary of terms in a book that provides a specific description of the meaning of terms used in the context of the book represents an *application ontology* well. As shown in Fig. 17.4, the levels of ontology are all related, and ontologies can be navigated from global to application and conversely, as well as at the same level.

In geographic information, the most common mechanisms used to develop ontologies are conceptual models, application schemas, feature catalogues, and dictionaries. In addition, typology and other classification methods may also be used.

17.5 Geosemantic Interoperability

Interoperability has been a subject of major development during recent decades. In the geographic information realm, it has been driven by standardization bodies, such as Technical Committee 211 of ISO (ISO/TC 211) and Open Geospatial Consortium (OGC), and also the research community. It provides solutions for sharing, integrating, and gaining access to geographic information from various sources. This section gives an account of interoperability of geographic information that is rooted in the communication process and provides a framework to integrate developments that encompass semantic interoperability of geographic information (i.e., geosemantic interoperability).

17.5.1 Communication: a Foundation for Geosemantic Interoperability

The study of communication began with *Shannon* [36, 37] and *Weiner* [38], to whom we attribute the initial development of *cybernetics* and information theory. Cybernetics is essentially concerned with the communication process and the controls that systems use to maintain order, organization, and balance. For instance, a driver applies control mechanisms constantly to maintain a car in the correct lane and the correct direction using the car's steering wheel, accelerator, and brakes. In a spatial database, the database manager models concepts, structures the data, and ensures consistency between data in order to facilitate data retrieval and use. Information is a notion directly related to order and balance. Information corresponds to the content that a system exchanges with others [38]. Good performance is dependent on information received and communication from/to others.

Information exchange between systems is essentially a question of efficient communication and control. Fundamentally, a communication process is composed of three components: a source, a message, and a destination ([36, 37, 39], Fig. 17.5).

A source can be an individual, a radio station, a geographic information system, a Web site, or any other system that can provide information. A source system's knowledge can only be communicated indirectly to another system. It is expressed by a message that is sent to a destination system. As is the case between human beings, the source selects the specific information to communicate, adapts it especially to the context of the destination, structures information elements, and finally encodes it as part of a message.

The message is essentially an ordered set of signs that may take various forms: ink on paper, sound waves in the air, light waves in an optical fiber, electrical power in a network (e.g., the Internet), etc. When a message is encoded, the source places it on the communication channel toward the destination. At this specific time, the message is released from the source and its initial attributed meaning. The source then no longer has control over the message. In the communication channel, the message has no intrinsic meaning and acts only as an intermediary or mediator between the source and its destination.

It is the destination, on receiving the message, that has the responsibility to first decode it and, second, assign it a meaning. As explained earlier based on the meaning triangle and cognitive functioning, the destination relies on its own knowledge to first decode the message and then to assign it a particular meaning.

Communication is only effective when the meaning of the message attributed by the destination corresponds to the source meaning at the time the message was placed in the communication channel.

There are a number of additional elements of communication that we have deliberately not included to keep this explanation simple. This is the case of the notion of *noise*, which adds disturbance to the message, and the notion of *feedback*, which is a control mechanism that sends messages back to the source in order to improve the communication between the source and the destination.

Individuals or systems engaged in a communication process are constantly working to maintain order in their exchanges with others in order to understand each other. It is because of their common knowledge, also called commonness ([39], Fig. 17.6), that the source and the destination can communicate.

If the destination has no common knowledge with the source, the communication cannot be effective and, therefore, the destination cannot decode and interpret the message properly. For example, two individuals must have common knowledge about the language they use to interact. If not, communication is not possible. Assuming they have common knowledge about the language, they must also share a certain number of similar concepts with respect to the vocabulary they use. For example, the term *bridge* can mean different things:

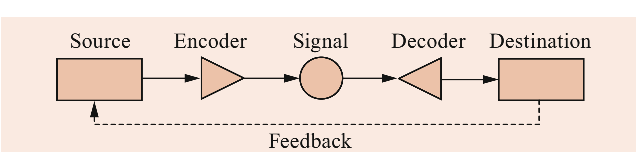


Fig. 17.5 The communication process

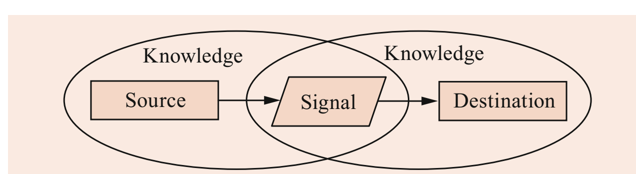


Fig. 17.6 Commonness

- A structure carrying a pathway or roadway over a depression or obstacle
- A piece raising the strings of a guitar or a violin
- The forward part of a ship's superstructure from which the ship is navigated
- A partial denture anchored to adjacent teeth.

If such knowledge is not common to both parties involved in a communication process, again, effective communication cannot be possible.

The notion of interoperability introduced by the computer science community is, in fact, very well aligned in pursuing the study of cybernetics and information theory. This is further demonstrated by the IEEE definition of interoperability as *the ability of two or more systems or components to exchange information and to use the information that has been exchanged* [35].

17.5.2 Heterogeneity in Geographic Information: a Barrier

Interoperability is essentially a question of efficient communication between systems. As one can recognize from the description of communication, it is concerned with both syntactic and semantic issues. However, geographic databases and users' representations of real-world phenomena show a number of heterogeneities between each other that interfere with efficient interoperability. Typically, heterogeneity has been decomposed into four levels: system, syntactic, structural, and semantic heterogeneity [28, 40, 41].

Databases often sit on different systems, and interoperability requires the establishment of an interconnection between them. With the development of communication networks and communication protocols such as Ethernet, Transmission Control Protocol/Internet Protocol (TCP/IP), Remote Procedure Call (RPC), File Transfer Protocol (FTP), Hypertext Transfer Protocol (HTTP), etc., the interconnection of systems working with various operating systems (Linux, Windows, Mac OS, etc.) is now possible. As such, databases of different types can also be interconnected to share data via applications such as Open Database Connectivity (ODBC), Java Database Connectivity (JDBC), and others.

Syntactic heterogeneity concerns the way data are physically represented [41], the signs used, and their order in a message. Syntax establishes the signs and the rules that define the order of the signs in a message. The form of the message is of primary importance here compared with its content in order to first be able to decode signs of the message. This relates to the data interchange format to deliver data (e.g., Geography Markup Language). Syntactic heterogeneity also refers to the various forms of data [42]; for example, geographic information can take the form of coverage, gridded data, and vector data.

Structural heterogeneity refers to differences in data modeling. In some databases, something may be described as a concept and in another as a characteristic of the concept. For example, *street* may be defined as a concept in one database, whereas it can be described as a characteristic of *road* in another one. The structural conflicts in geographic information can be classified as follows: concept, property, geometry, and temporality.

Semantic heterogeneity refers to differences of meaning between concepts. As shown in Fig. 17.1, a concept's meaning results from the link made with the sign and the object. The difference in cognitive models between individuals that associate identical signs with different phenomena and the same phenomena with different signs illustrates the problem of semantic heterogeneity. Context plays an essential role since cognitive models are developed in specific contexts. In fact, it is the context that provides semantics to real-world phenomena; it guides how phenomena are perceived and abstracted, and then influences their definition in terms of concepts. It is essential to take the context into consideration when reasoning, for instance, when resolving semantic heterogeneity and similarity between concepts. Context provides details on:

Use:	user identification, user profile, user location, type of uses
Data:	source, geospatial entities, meaning, scale, date of validity, etc.
Association:	relationships (spatial, semantic, etc.)
Procedure:	process steps to capture the data, query to get the data, etc.

Metadata constitute a valuable source of contextual details.

17.5.3 Geosemantic Interoperability Depicted from a Communication Perspective

Geographic information interoperability can be compared to an interpersonal communication process. Let us describe it in the context where two agents, i.e., a user (A_u) and a data provider (A_p), interact together regarding geographic information (Fig. 17.7); A_u would like to get information about some geographic features—say the road network, in a given region—say Sherbrooke, for delivery purposes; A_u sends a request for road information to A_p over the Internet (i.e., the communication channel) in its own vocabulary. As soon as the request reaches A_p, A_p interprets it in order to identify and gather the information it has that answers the request of A_u (e.g., King Street, Portland Blvd, etc.;

17

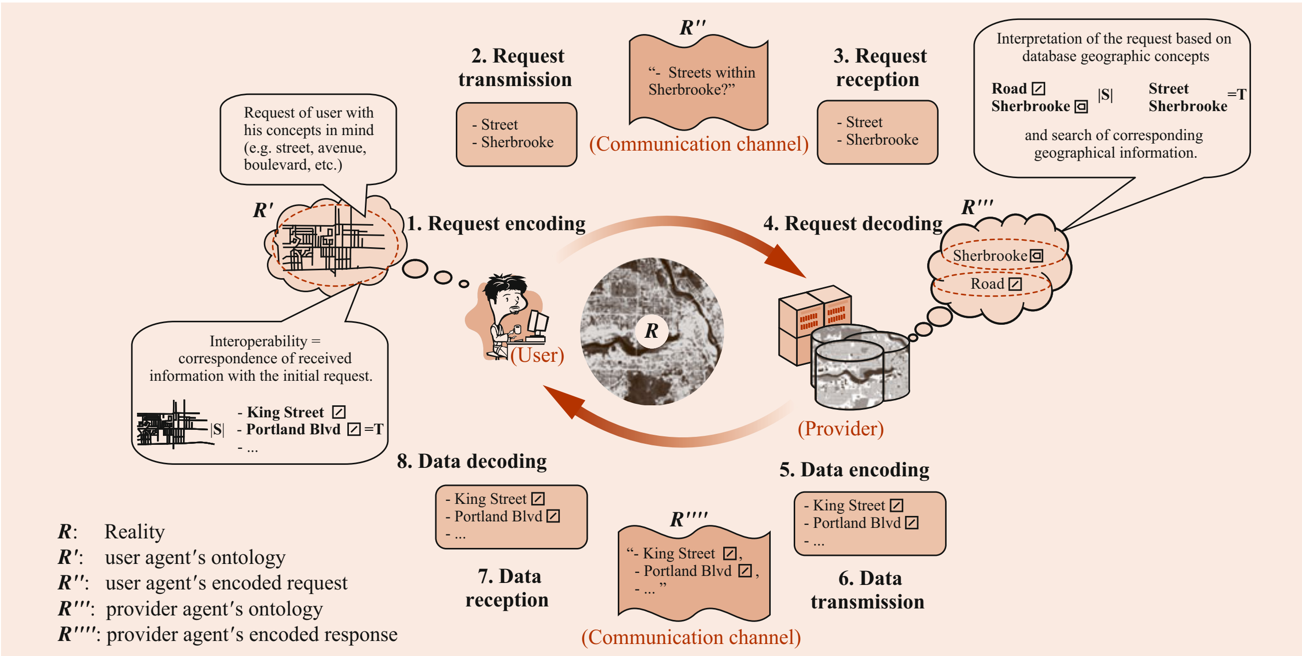


Fig. 17.7 Framework for geosemantic interoperability (S = similarity operator, T = true)

the spatial pictograms indicate one-dimensional (1-D) and two-dimensional (2-D) objects [43]). In turn, when A_u gets a response back about its request from A_p, it interprets it in order to determine if the response fulfills the request. There is interoperability between the two agents if the initial request is satisfied.

In this situation, the two agents are using their own vocabularies to express abstractions of real-world phenomena. As long as agents have a common background and set of symbols, they usually end up understanding each other.

This situation also exhibits five different expressions of reality, denoted by R, R', R'', R''', and R''''. These expressions are related to each other as part of the communication process. First of all, there is the road network of the topographic reality (R) as it exists at a given time and appears to A_u. This is the A_u's referent.

Second, A_u has developed its own cognitive representation of the road network, i.e., R', from its observation of it and its reference frame, i.e., the set of rules and knowledge its uses to abstract phenomena. It is made of significant properties (e.g., geometric, temporal, and descriptive properties, behavior, and relationships) that are joined and structured together in concepts (i.e., signified). Consequently, a concept is merely a simplified and fictional representation of a real-world phenomenon or a part of it [44], which corresponds to the perceptual symbol that an agent has in mind [11, 39, 45]; R' consists in A_u's ontology.

Since concepts are only theoretical representations of real-world phenomena, A_u cannot communicate them directly. Consequently, concepts must be transformed into physical forms (e.g., sounds, bits and bytes, etc.) to compose a message. This is the encoding operation. By this process, A_u transforms concepts of the road network into signs, such as words, abbreviations, punctuation, symbols, and pictograms, organized in a specific manner such as "street within Sherbrooke." This is denoted by R'' and refers to the signifiers. Once encoded, the message can be released on the communication channel toward A_p. Once released, the message loses its meaning and becomes the data transmitted and used for the purpose of interoperability.

When A_p receives the message, it begins the decoding operation that is essential for message interpretation. As such, A_p relies on its own set of concepts, i.e., R''', to assign a proper meaning to the message. In our situation, this corresponds to Road and Sherbrooke. Here, R''' serves as A_p's ontology.

At the time A_p has attributed a meaning to the message, it initiates the retrieval and gathering of information that fulfills the request (i.e., King Street, Portland Blvd, etc.). Then, this information is encoded and placed on the communication channel as data toward A_u. This corresponds to R''''.

In the end, when A_u gets the response to their request, it initiates the message decoding, identifies the concepts it knows to give a meaning to the message, and, finally, assesses whether the response answers its initial request correctly and infers the concepts that constituted the requests. When this is the case, interoperability has occurred between the two agents.

Interoperability is presented here as a bidirectional mechanism as opposed to a pipeline processing information in

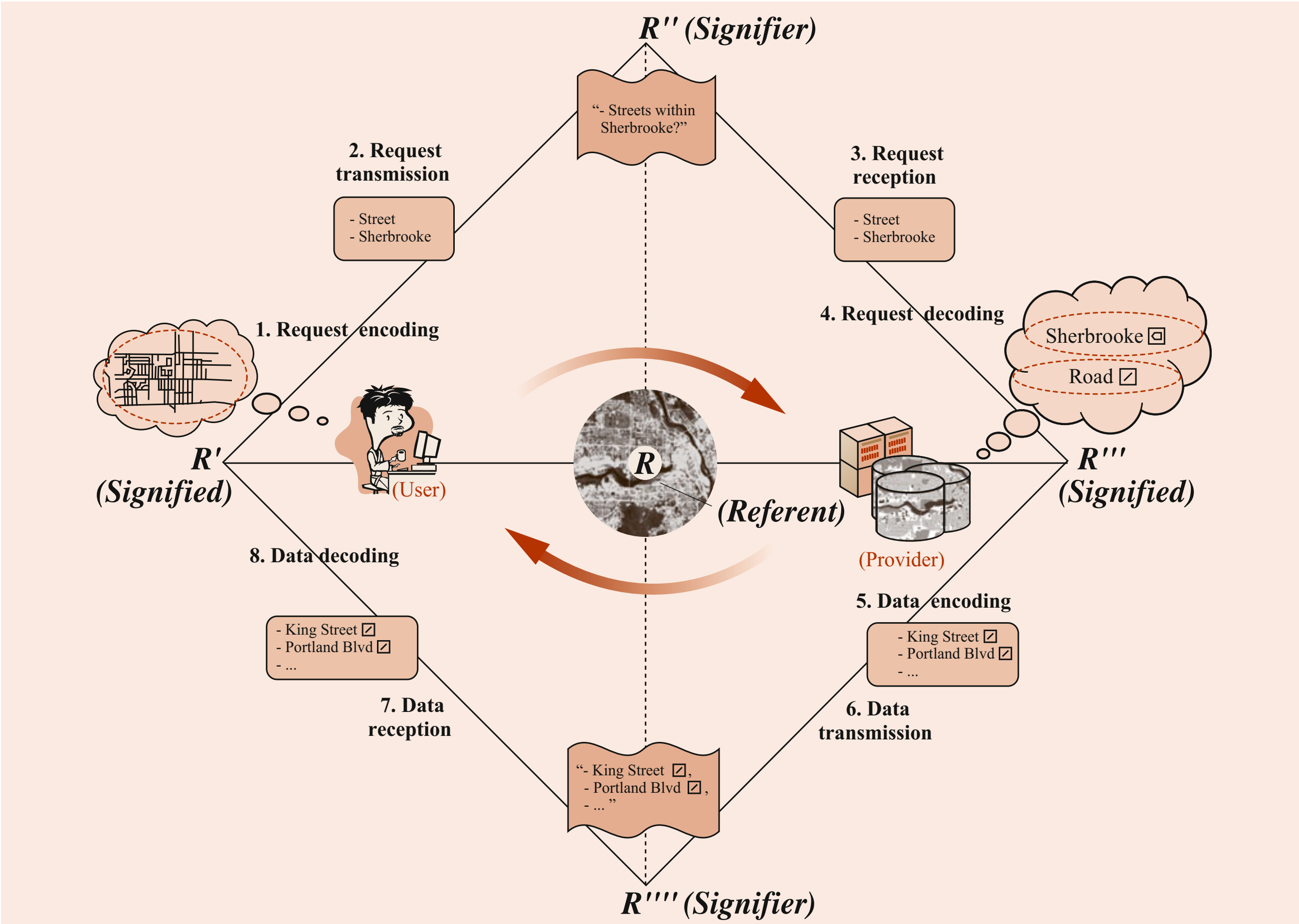


Fig. 17.8 Semiotics and geosemantic interoperability

one direction only (source to destination). This goes beyond the simple capability to gain access to information from geographic databases and display it on a screen or print it on paper. Most of the time, the exact vocabulary of geographic databases must be known in advance in order to obtain the appropriate information. However, interoperability means that users and providers must have the capability to semantically understand queries and responses. In the context of the Web and the Semantic Web, such knowledge and capability are increasingly becoming available.

17.5.4 The Importance of Geosemantic Interoperability

The framework for geosemantic interoperability (Fig. 17.7) is composed of four distinct instances of the semiotic triangle (Fig. 17.1), which are specifically depicted in Fig. 17.8. The first one arises between the reality (R), A_u's ontology (R'), and the encoded request (R''). The second is between the reality (R), the encoded request (R''), and A_p's ontology (R'''). The third one appears between the reality (R), A_p's ontology (R'''), and the encoded response (R''''). Finally, the fourth instance is between the reality (R), the encoded response (R''''), and A_u's ontology (R').

This illustrates how semantics is deeply embedded in the framework for geosemantic interoperability and, thus, provides a comprehensive account of semantic interoperability as a whole that underlies the development of semantic spatial data infrastructure and furthermore the Geosemantic Web.

17.6 Spatial Data Infrastructure and Geospatial Information Environment in Support of Geosemantic Interoperability

Geosemantic interoperability, as explained in the previous section, constitutes a foundation for the development and implementation of spatial data infrastructures (SDIs) and geospatial information environments [46]. This section aims to explain how geosemantic interoperability is basic to SDIs and geospatial information environments.

17.6.1 Spatial Data Infrastructure

The purpose of an SDI is to facilitate and coordinate the useful exchange and sharing of geographic information and services in a community [46]. It is the most common means by which governments and organizations have developed interoperability for geographic information. An SDI is basically composed of five elements, i.e. policies, technologies, standards, human resources, and associated activities required for collection, processing, management, access, delivery, and use of geographic information [47, 48]. SDIs have mostly been developed on the basis of the Reference model for Open Distributed Processing (RM-ODP) as described in ISO/IEC 10746-1:1998 [49]. RM-ODP consists of a framework of five viewpoints: the enterprise viewpoint, the information viewpoint, the computational viewpoint, the engineering viewpoint, and the technology viewpoint. Briefly, in the context of an SDI, the enterprise viewpoint addresses the purpose and the scope, the policies, the responsibilities, and the business process of the SDI. It defines the role of the SDI in its environment. The information viewpoint focuses essentially on the information made available through the SDI and its semantics, which is essential for geosemantic interoperability. The computational viewpoint concerns the functional SDI decomposition into services with interfaces and operations. This viewpoint is of great interest for the definition of semantic components and services. The engineering viewpoint is mainly related to the interaction between data and services, and system interconnections. Finally, the technology viewpoint (Chap. 15) refers to the specifically chosen technology for the SDI implementation.

As one can deduce from the above five viewpoints, the first three have a special interest and importance with respect to semantic interoperability of geographic information. It is in the enterprise viewpoint that the goal of achieving semantic interoperability must be clearly stated. Additionally, the enterprise viewpoint shall identify any stakeholder that must participate in the SDI, which would include, for instance, brokers that will assist users and providers in finding appropriate data with respect to their respective specific vocabulary and semantics. As explained in the previous section, it is not humanly conceivable that all users can know in advance the exact vocabulary and semantics used by the multiple geographic information sources accessible in an SDI. Users should be able to interact with the SDI in their own vocabulary and find data that fits their specific purpose. Consequently, the information viewpoint must include the essential informational components to enable semantic queries (i.e., queries made in the user's vocabulary and correctly interpreted by the information server). Ontologies must, therefore, be integrated as part of the information viewpoint to provide the fundamental knowledge for reasoning, query interpretation, and appropriate responses. The computational viewpoint also needs specific attention to address semantic issues in defining semantic interfaces linked with ontologies, including reasoning operations and functions that assist the interpretation of queries and responses. This viewpoint also includes the encoding of ontologies to interface with semantic services.

Addressing semantics early in the definition of an SDI facilitates its design and allows clear identification of the ontologies and semantic services that are required.

17.6.2 Geospatial Information Environment

A geospatial information environment is the basic notion that fostered the development of SDIs. The notion of "information environment" was defined in [50] as an "aggregate of individuals, organizations, and systems that collect, process, disseminate, or act on information, also included is the information itself." This notion was further extended in the context of geospatial information to add all aspects closely related to spatial information, namely the data structure and the system interfaces for data dissemination [46].

Essentially, a geospatial information environment consists of a set of resources that enables access to geospatial information with its understanding. It covers the following components: data structure/schema (application schema, [51]), data description/semantics (feature catalogue, [52]), metadata, data and metadata capture operations, data (the data elements), data management, discovery, access, and transformation (Fig. 17.9). These components are thoroughly addressed by ISO (ISO/TC 211) and OGC geographic information standards.

On the one hand, in a geospatial information environment, there are users who access the system by the way of information services. They may want to discover by using cataloguing services (e.g., OGC CSW) what sort of data is available from the geospatial information environment, get access to it via various types of services (e.g., OGC WMS [53], WFS [54], WCS [55], WPS [56]) for different purposes, or get direct access to it by the way of geographic information system (GIS) for data generation and management purposes. On the other hand, there is geospatial data, which is typically defined, structured, and organized in terms of feature types, properties (thematic, spatial, temporal, metadata), and associations between feature types [51] that are documented as part of application schemas, feature catalogues [52], and ontologies. However, these elements may be crafted differently from one geospatial data environment to another. In the geospatial data store, the application schemas, feature catalogues, and ontologies are maintained in a schema repository for the geospatial information envi-

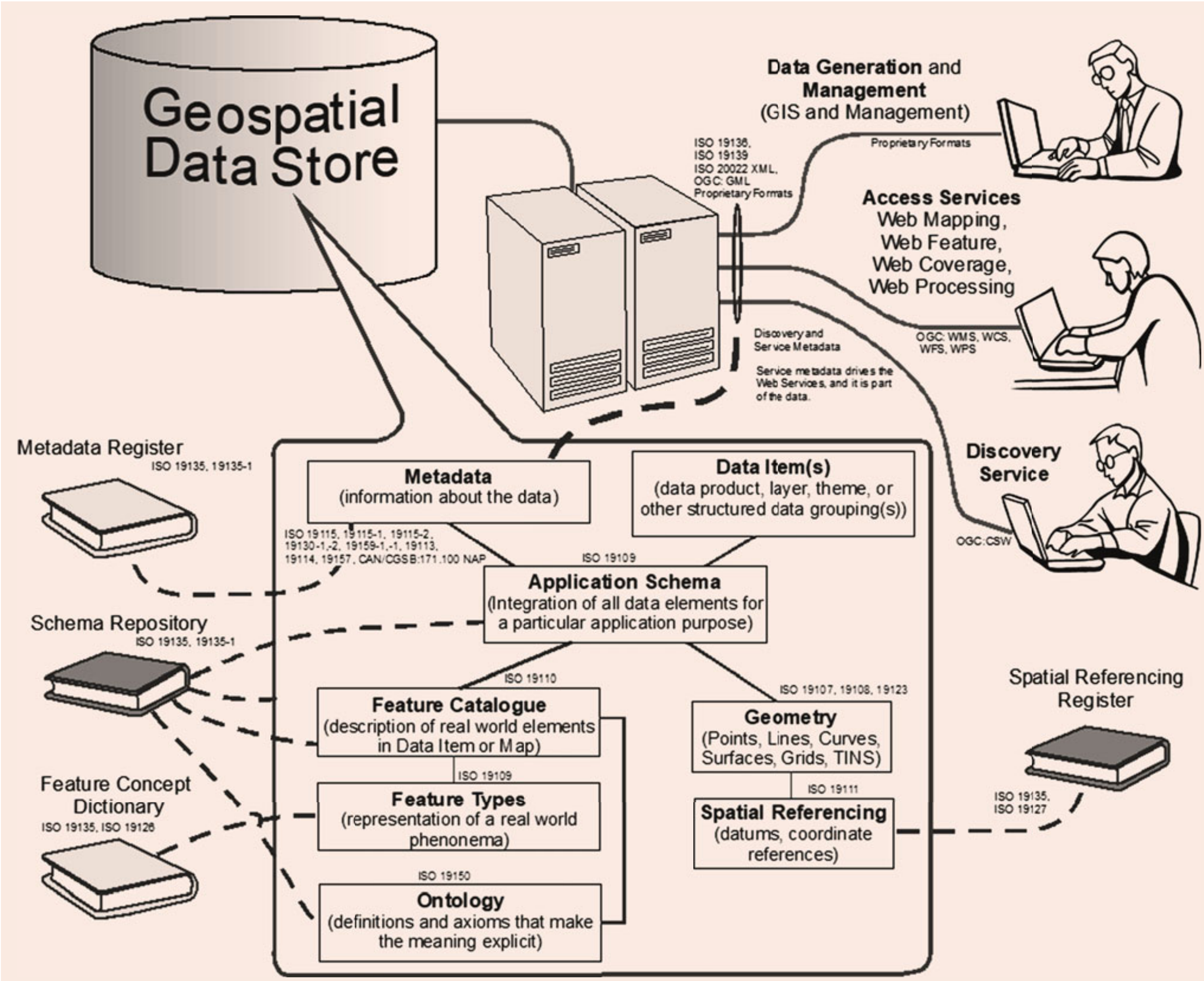


Fig. 17.9 Geospatial information environment (© Crown)

ronment. Therefore, it is fundamental for users and systems to get access to such information, which provides the essential semantic description, to gain the knowledge required for interpretation and understanding of the geospatial data maintained and provided by the geospatial data store, and consequently enable geosemantic interoperability.

In addition, a geospatial information environment provides geospatial metadata, which provides further details about the geospatial data, such as its temporal validity, spatial accuracy, data acquisition process, and more. A spatial referencing register can be integrated or referenced in the geospatial information environment to catalogue the geodetic codes and parameters that are used.

The deployment of a geospatial information environment is essentially oriented toward supporting the development of SDIs that are geosemantic interoperability enabled.

17.7 Geographic Information Standards – a Key Element for Geosemantic Interoperability

As introduced previously, ISO/TC 211 [57] is working on the development of a suite of standards (known as the ISO geographic information suite of standards) to enable interoperability of geographic information. These standards are required to support the understanding and usage of geographic information, to increase the availability, access, integration, and sharing of geographic information, and to advance the development of SDIs and geospatial information environments. It provides the fundamental structure and semantics for the description and representation of geographic information.

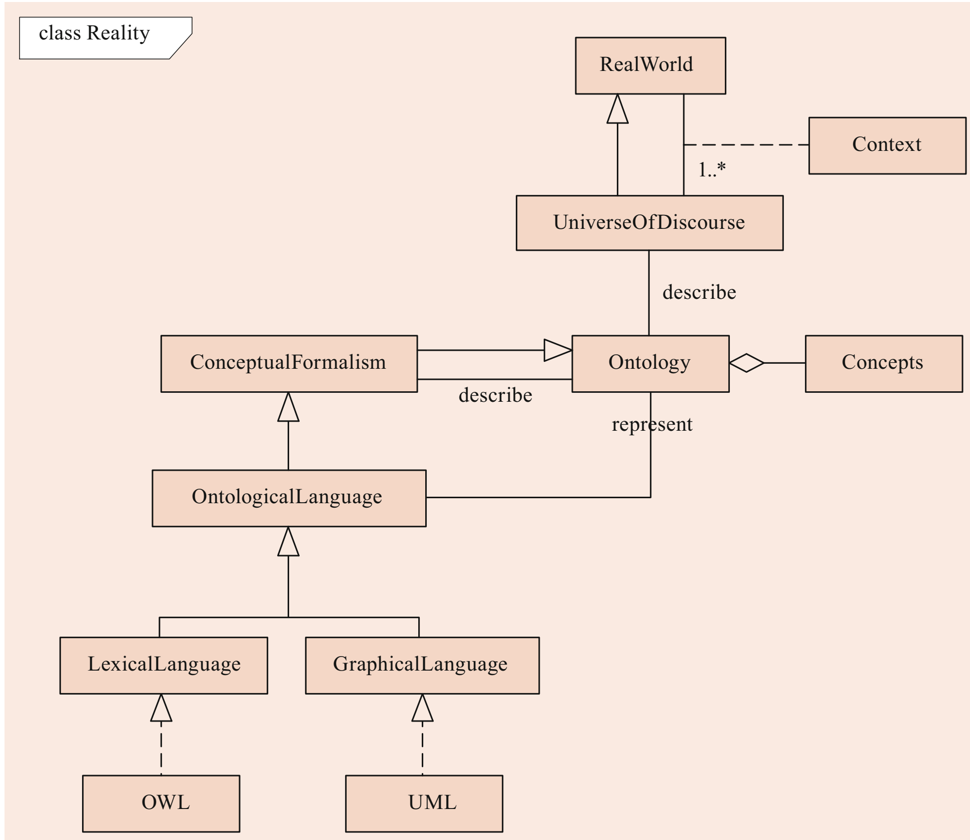


Fig. 17.10 ISO/TC 211 Real-world abstraction description and ontological languages

An RM-ODP (Sect. 17.6.1) -based reference model [58] sets the foundation of the suite of standards. Among others, semantics of geographic information has been recognized as one of the three foundations for the development of geographic information standards. This foundation is meant to provide knowledge about geographic information, which, therefore, allows reasoning. The other two foundations are the syntactic and service foundations.

The semantic foundation is concerned with the abstraction of geographic phenomena with the meaning and the structure of defined geographic concepts, including thematic, spatial, and temporal characteristics and their documentation. Thus, concepts covering geometry, topology, temporality, spatial referencing (either directly through coordinates, or more indirectly by use of, for instance, area codes like postal or zip codes, addresses, etc.), any information providing context (e.g., quality, metadata) are also of matters in this foundation. In ISO/TC 211, the semantic foundation also includes the definition of top-level ontologies, which make ontology mapping between specific domains possible.

The Unified Modeling Language (UML) [59] and the Ontology Web Language (OWL) [60] are recognized as the two languages for concept definition and description (graphical and lexical, respectively; Fig. 17.10).

The ISO technical specification ISO/TS 19150-1, Geographic information – Ontology – Part 1: Framework [61], describes the manner that ISO/TC 211 is developing semantic resources about geographic information to promote and facilitate the use of ISO geographic information standards for the Semantic Web (Fig. 17.11).

ISO 19150-2, Geographic information – Ontology – Part 2: Rules for developing ontologies in the Web Ontology Language (OWL) [62], sets rules and guidelines for the derivation of the UML static view modeling elements used in the ISO geographic information standards and ISO 19109, Geographic information – Rules for application schema-based [51] application schemas into OWL.

ISO19150-3, Geographic information – Ontology – Part 3: Semantic operators [61], aims to set semantic proximity operators between concepts associated with geometric and temporal representations. These operators should complement the current suites of geometric and temporal operators already defined in ISO 19107 [63], ISO 19108 [64], ISO 19125-1 [65], and ISO 19141 [66].

ISO19150-4, Geographic information – Ontology – Part 4: Service ontology [67], sets a framework for geographic information service ontology and the description of geographic information Web services in OWL.

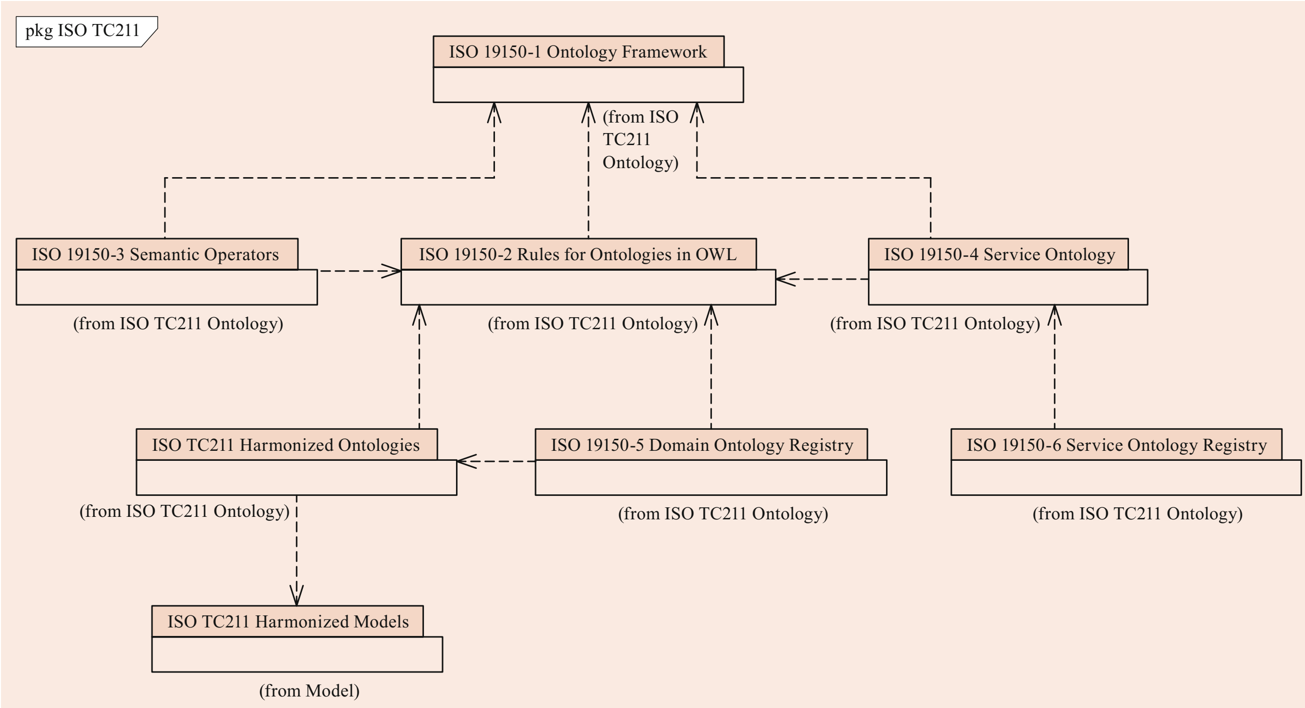


Fig. 17.11 ISO/TC 211 Resources on ISO geographic information standards for the Semantic Web

ISO19150-5, Geographic information – Ontology – Part 5: Domain ontology registry [61], aims to set an international registry for geographic information domain ontologies and its maintenance. The registry is meant for ISO standardized high level geographic information ontologies.

ISO19150-6, Geographic information – Ontology – Part 6: Service ontology registry [61], aims to set an international registry for geographic information service ontologies and its maintenance. The registry is meant for standardized geographic information service ontologies.

The ISO/TC 211 Harmonized ontologies consists of a suite of OWL ontologies derived, in compliance with ISO 19150-2, from the UML models of the suite of ISO geographic information standards as maintained in the ISO/TC 211 Harmonized model [68]. The ontologies can be found on the Web at https://github.com/ISO-TC211/GOM/tree/master/isotc211_GOM_harmonizedOntology. The ISO/TC 211 Group for Ontology Maintenance (GOM) ensure the derivation and the maintenance of the ontologies. The ontologies are derived automatically using a set of JavaScripts within enterprise architect (SPARX Systems®), a UML modeling software.

17.8 Geospatial Semantic Web Aiming at Geosemantic Interoperability

Early in the new millennium, the Web progressed significantly toward the Semantic Web (Web 3.0). As you can experience when surfing the Web, it is becoming a tremendous open worldwide database [4]. The Web can now be queried from a people-specific perspective, understanding, and perception of real-world phenomena to get accurate, detailed, and appropriate answers. Web browsers are becoming increasingly efficient, adopting Semantic Web elements and capabilities (Resource Description Framework (RDF) tagging, reasoning capabilities with ontologies, etc.).

17.8.1 A Brief Description of the Semantic Web

In 2001, *Berners-Lee* advanced the idea of the Semantic Web [69]. His main conviction was to turn the Web from a Web of documents perspective to a Web of data and information perspective. On the Semantic Web, data must be understandable and processable by computers and, therefore, the Semantic Web would be capable of answering questions as opposed to solely returning documents or Web pages corresponding to certain keyword criteria. The Semantic Web is a Web of application-independent data data that can be composed from multiple sources, data that is arranged into class systems, data that is part of a larger information system by way of ontologies [24]. Data in the Semantic Web is much smarter compared with the Web. At the lowest level, data may take the form of documents and records in spreadsheets or databases. A smarter level would be Extensible Markup Language (XML) documents with associated vocabularies. At the upper level, data are coupled with ontologies and reasoning capabilities to infer new knowledge (Fig. 17.12).

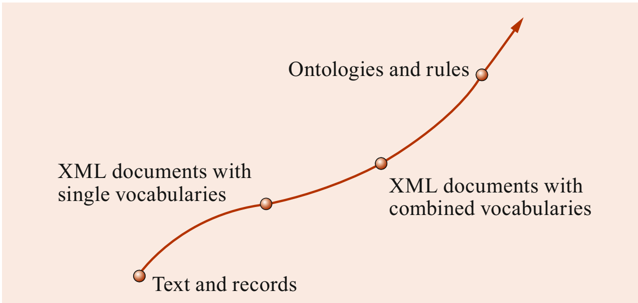


Fig. 17.12 A progression towards clever data (adapted from [24])

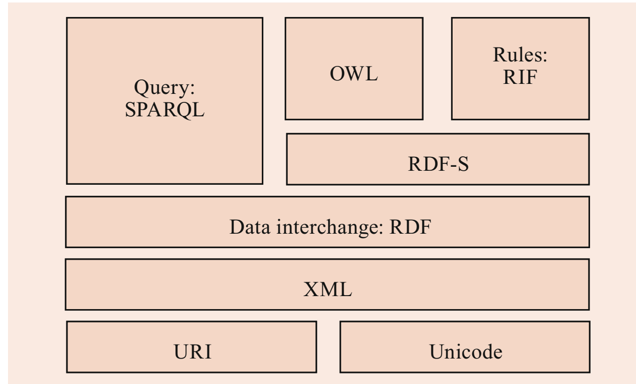


Fig. 17.13 The Semantic Web foundation (extracted from the Semantic Web stack in [70])

Accordingly, the Semantic Web is about interoperability of data over the Web [70]. According to *Daconta* et al. [24], the Semantic Web consists in a solution for

1. Information overload, especially with the propagation of the Internet
2. Breaking stovepipe systems and allowing information sharing
3. Integrating and aggregating data from various sources on the Web
4. Enabling more efficient data retrieval considering users' and data sources' specific vocabularies (or concepts).

As one can recognize, ontology, therefore, underpins the Semantic Web. It defines the meaning of data and describes it in a format that is computer and application readable and processable. Thus, when applications use data, they can consider data's inherent semantics at the same time, making them more skillful and accurate. This additional capability facilitates the integration of heterogeneous data captured by various communities based on the similarity of the semantics.

The Semantic Web requires logical assertions, concept classification, formal models, rules, and trusts. Logical assertions allow the association of a subject to an object through a verb. For example, in the assertion "Natural Resources Canada is the author of CanVec" [71], the subject "Natural Resources Canada" is associated to the object "CanVec" via the verb "is the author." Concept classification developed as ontologies provides the required vocabulary and semantics. Formal models are necessary to set the structure for the definition of concepts with their properties and relations (e.g., the ISO/TC 211 General Feature Model) to support reasoning. Rules are required for the derivation of conclusions by inference. For example, if the CanVec map 21E05 was revised as of August 1, 2006, and the CanVec map 21E06 was revised as of October 1, 2008, we can then infer that map CanVec 21E06 is more recent than CanVec map 21E05. Trust is a notion related to the reliability of data sources and reassures users about the confidence of data. The notion of trust is also related to digital signatures for certification.

The architecture of the Semantic Web has been widely described [70] and is summarized briefly hereafter. As illustrated in Fig. 17.13, it is founded on uniform resource identifiers (URI), the universal character set (Unicode), and the Extensible Markup Language (XML).

URIs are strings of characters that serve as unique identifiers to designate a resource, either physical or abstract, on the Web. A resource can be about things, persons, and concepts. URIs have a specific syntax that follows the standards called RFC 3986 [72], which is available at [72]. URIs include both uniform resource locators (URLs) and uniform resource names (URNs). The URL identifies the place where the resource is available on the Internet and how to get to it. An example of a URL is http://www.isotc211.org/. A URN is typically used for identification purposes. An example of a URN is *urn:iso:19115:-1:ed-1:en*.

Unicode is an industry standard that is used as the basis for text and language for the Semantic Web. It allows the representation and manipulation of text strings in diverse languages through the standardization of a character repertoire and character codes that comply with ISO standard 10646 Information technology – Universal Coded Character Set (UCS) [73]. A unique number specifies each character independently of the system, application, and language. Accordingly, Unicode allows systems and applications to exchange character-based information and to browse Web page information without loss of information.

XML is the syntactic foundation for the Semantic Web. On this basis, a stack of XML syntaxes and vocabulary has been elaborated by the World Wide Web Consortium [74]: RDF, RDF-S, Web Ontology Language (OWL), SPARQL protocol and RDF query language (SPARQL), and Rules Interchange Format (RIF). The first was the Resource Description Framework (RDF). RDF enables the representation of information about specific resources or things that exist on the Web. It uses a triple structure commonly called the subject–predicate–object triple (Fig. 17.14). The subject is the

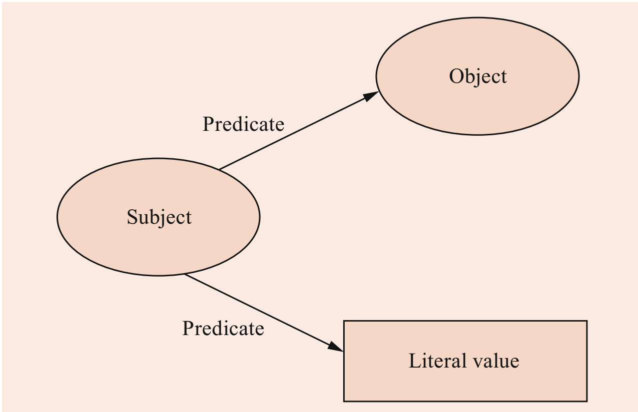


Fig. 17.14 The subject–predicate–object triple

thing about which something is asserted. The predicate is the relation that binds the subject to the object. The object is either a literal value describing the subject or another resource referred to the subject by the predicate.

For example, the assertion "Natural Resources Canada is the author of the map 21E05" may be represented in RDF as such

```
< rdf: Description
  rdf: about ="http:www.nrcan.gc.ca">
< ex: isAuthor>
< rdf: Description
  rdf: about="http://open.canada.ca/
       data/en/dataset/
       8ba2aa2a-7bb9-4448-b4d7-f164409fe056"/>
< /ex: sAuthor>
< /rdf: Description> ,
```

where http://www.nrcan.gc.ca is a Web resource corresponding to Natural Resources Canada, isAuthor is the predicate, and http://open.canada.ca/data/en/dataset/8ba2aa2a-7bb9-4448-b4d7-f164409fe056 is the Web resource corresponding to CanVec.

RDF schema (RDF-S) defines classes and properties required for the description of classes, properties, and other resources. It is based on RDF. RDF-S resources serve for the creation of application or user community-specific RDF vocabularies and, furthermore, for creating classes for specific data. Class instances can then be defined in RDF.

SPARQL, which stands for SPARQL Protocol and RDF Query Language, is the Semantic Web language allowing interrogation of RDF data sources and OWL ontologies. As such, the structure of the language and the query processor aims at analyzing RDF triples to identify those matching the query.

The Web Ontology Language (OWL) was developed for the representation of knowledge. It consists of an evolution of the existing DAML+OIL (DAML: DARPA Agent Markup Language; OIL: Ontology Inference Layer). It is based on RDF and RDF-S. There are various levels of expressivity or profiles in OWL: OWL1 Lite, OWL1 DL, OWL1 Full, OWL2 DL, OWL2 EL, OWL2 QL, OWL2 RL, and OWL2 Full. It is not the goal here to describe the details, but briefly:

- OWL1 Lite: It is intended mainly for the description of classification hierarchy with attributes. Cardinalities are limited to 0 or 1.
- OWL1 DL: DL stands for Description Logics. This adds knowledge representation that improves reasoning. More flexibility is allowed on cardinality restrictions.
- OWL1 Full: This allows maximum expressiveness and the syntactic freedom of RDF(S). The document must be RDF(S) valid.
- OWL2 DL: It extends OWL1 DL with additional possibilities about type separation, punning, declaration, disjoint classes, role characteristics, etc.
- OWL2 EL: This is a lightweight description logic language that fits more specifically ontologies that are composed of a very large number of classes and properties while using a limited number of OWL functionalities.
- OWL2 QL: It is meant to support data-driven applications and conjunctive query answering implementable in relational database systems.
- OWL2 RL: It is meant to support rule-based reasoning and inference technologies.
- OWL Full: This is the aggregation of OWL2 DL and RDF(S).

Figure 17.15 illustrates a snippet of OWL for the definition of the ISO 19115-1 classes CI_Citation, CI_Organization, CI_Party, and CI_Responsibility, where CI_Organization is as a subclass of CI_Party.

OWL was specifically chosen by ISO/TC 211 for the semantic representation of the concepts defined in its geographic information standards to support the Semantic Web. OWL is a fundamental piece of the Semantic Web for carrying out reasoning, interpretation, and inference. Ontologies support software agents for the interpretation of incoming data elements. Software agents can automatically assign a meaning to an incoming message (query or data). Such a function is made possible by ontology matching or similarity assessment.

Ontology matching refers to the correspondence between semantically related entities of distinct ontologies [75]. It aims at finding concepts and/or properties from different ontologies that are related together, such as equivalence, subsumption, and disjointness. Mapping of concepts and properties between ontologies is not a trivial function. Typ-

17

```xml
<!--+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++-->
<!-- File name:          iso19115-1CitationAndResponsiblePartyInformationExtract.rdf -->
<!-- Generator:          GOM_Technologies::UML2OWL                                   -->
<!-- Generator version:  2.5                                                         -->
<!-- File date:          9-10-2015                                                   -->
<!--+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++-->

<rdf:RDF xmlns="http://def.isotc211.org/iso19115/-1/2014/CitationAndResponsiblePartyInformation#"
  xml:base="http://def.isotc211.org/iso19115/-1/2014/CitationAndResponsiblePartyInformation"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">

  <!--+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++-->
  <!-- Ontology: ISO 19115-1:2014 Citation and responsible party information -->
  <!--+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++-->
  <owl:Ontology rdf:about="http://def.isotc211.org/iso19115/-1/2014/CitationAndResponsiblePartyInformation">
    <rdfs:label>ISO 19115-1:2014 Citation and responsible party information</rdfs:label>
    <owl:versionInfo>2014</owl:versionInfo>
  </owl:Ontology>

  <!--+++++++++++++++++++++-->
  <!-- Class: CI_Citation -->
  <!--+++++++++++++++++++++-->
  <owl:Class rdf:about="&iso19115-1;CI_Citation">
    <rdfs:label>CI_Citation</rdfs:label>
    <skos:definition>standardized resource reference</skos:definition>
    <rdfs:isDefinedBy>http://standards.iso.org/iso/19115/-1/ed-1/en/</rdfs:isDefinedBy>
  </owl:Class>

  <!--+++++++++++++++++++++++++-->
  <!-- Class: CI_Organisation -->
  <!--+++++++++++++++++++++++++-->
  <owl:Class rdf:about="&iso19115-1;CI_Organisation">
    <rdfs:label>CI_Organisation</rdfs:label>
    <skos:definition>information about the party if the party is an organisation</skos:definition>
    <rdfs:isDefinedBy>http://standards.iso.org/iso/19115/-1/ed-1/en/</rdfs:isDefinedBy>
    <iso19150-2:constraint>count (name + logo) .gt. 0</iso19150-2:constraint>
    <rdfs:subClassOf rdf:resource="&iso19115-1;CI_Party"/>
  </owl:Class>

  <!--++++++++++++++++++-->
  <!-- Class: CI_Party -->
  <!--++++++++++++++++++-->
  <owl:Class rdf:about="&iso19115-1;CI_Party">
    <rdfs:label>CI_Party</rdfs:label>
    <skos:definition>information about the individual and.sl.or organisation of the party</skos:definition>
    <rdfs:isDefinedBy>http://standards.iso.org/iso/19115/-1/ed-1/en/</rdfs:isDefinedBy>
    <iso19150-2:isAbstract rdf:datatype="&xsd;boolean">true</iso19150-2:isAbstract>
  </owl:Class>

  <!--+++++++++++++++++++++++++++-->
  <!-- Class: CI_Responsibility -->
  <!--+++++++++++++++++++++++++++-->
  <owl:Class rdf:about="&iso19115-1;CI_Responsibility">
    <rdfs:label>CI_Responsibility</rdfs:label>
    <skos:definition>information about the party and their role</skos:definition>
    <rdfs:isDefinedBy>http://standards.iso.org/iso/19115/-1/ed-1/en/</rdfs:isDefinedBy>
  </owl:Class>

</rdf:RDF>
```

Fig. 17.15 Example of ISO 19115-1 classes in OWL

ically, one or more concepts from one ontology can map to one concept of another ontology, and conversely. A similarity assessment between concepts supports the mapping between concepts. The similarity assessment can be either quantitative or qualitative. Typically, the quantitative assessment similarity takes the form of a semantic distance between concepts and is expressed in the interval [0, 1], where 0 means that the concepts are disjoint, and 1 means that the concepts are equivalent. Qualitative similarity assessment takes the form of predicates that specify the relation between two concepts, such as equal, include, overlap, etc.

In the spatial domain, the similarity between geometric constructs (point, curve, surface, and solid) is commonly identified by an intersection matrix between the interior, boundary, and exterior of two geometric objects, which leads to a 9 intersection matrix [63, 76]. Although quantitative assessment of semantic similarity is interesting, qualitative assessment of semantic similarity is closer to cognitive reasoning. An approach that follows the same paradigm as for the geometric similarity assessment (9 intersection matrix) would have the benefit of expressing the similarity in the same manner using the same type of operators. The geosemantic proximity [77] approach is a 4 intersection matrix assessment of the semantic similarity between contexts of two geospatial concepts, which compares the intrinsic and extrinsic properties between two concepts. The context of

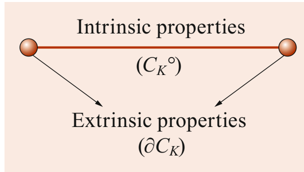


Fig. 17.16 The segment metaphor for concept's context

a concept can be represented using a geometric line metaphor (Fig. 17.16) where the intrinsic properties of the concept are compared with the interior of the line and the extrinsic properties of the concept, to the boundary of the line.

Then, the geosemantic proximity (GsP) between two concepts' contexts is expressed through the intersection of the intrinsic and extrinsic properties (Fig. 17.17), and 16 (2^4) predicates are then defined.

Each matrix component is evaluated whether or not it is empty (denoted by $\varnothing/\neg\varnothing$ or f/t, respectively) and, hence, 16 geosemantic proximity predicates are derived. The predicates are structured beginning with the acronym "GsP" followed by an underscore character "_" and the results of the four intersections of the matrix presented in row major form (i.e., row by row). The 16 predicates are GsP_ffff (disjoint), GsP_ffft, GsP_fftt (contains), GsP_tfft (equal), GsP_ftft (inside), GsP_tftt (covers), GsP_ttft (coveredBy), GsP_fttt (overlap), GsP_tttt, GsP_tfff (meet), GsP_tftf, GsP_tttf, GsP_ttff, GsP_fttf, GsP_fftf, and GsP_ftff.

Reasoning is the process by which one draws inferences or conclusions. Reasoning and inference are made possible through the relations existing between concepts. Such relations can be of the type:

Subsumption: the semantic relation that incorporates something under a more general category, e.g., isA, isSuperclassOf relationships

Meronymy: the semantic relation that holds between a part and the whole, e.g., partOf, hasA relationships

Semantic similarity: the semantic relation that expresses the closeness between concepts, e.g., geosemantic proximity or matching distance.

Examples

IF ontologyA:street $\subseteq$ ontologyA:road AND ontologyA:road = ontologyB:thoroughfare THEN ontologyA:street $\subseteq$ ontologyB:thoroughfare

IF ontologyA:watercourse contains (or GsP fftt) ontologyB:river/stream AND ontologyB:river/stream contains (or GsP fftt) ontologyC:creek THEN ontologyA:watercourse contains (or GsP fftt) ontologyC:creek

IF Joe is passenger of Train 1234 AND Train 1234 goes to Rome THEN Joe goes to Rome

Because ontologies provide the definitions of concepts and relationships between them with respect to some context, reasoning and inference are then possible. Reasoning and inference can lead to a wide range of possibilities related to interoperability: data discovery, query answering, composition of geographic data from multiple sources, using data across domains, data integration and mash up, and so on.

The Rules Interchange Format (RIF) refers to a series of World Wide Web formats to define various kinds of rule systems to support the interoperability among rule languages in general but also the interchange of rules in rule-based systems on the Semantic Web.

The Semantic Web has been described from a data perspective. However, the current Web is also composed of Web services. Currently, the use and interaction of Web services still require the participation of humans to find and integrate Web services. The Semantic Web could also contribute to facilitate the interaction with Web services by introducing semantics in Web services [4]. The Semantic Web can support Web services to automate Web Services Discovery, the composition of Web services, and their invocation in order to enable seamless machine interoperation with minimum human interaction. Semantic annotation of services in terms of capabilities, selection, access, composition, and invocation are then required and should be supported through ontologies.

A number of frameworks have been proposed to support the above functionalities (e.g., OWL-S-Semantic Markup for Web Services [78], Semantic Web Services Ontology (SWSO) [79], and Web Service Modeling Ontology (WSMO) [80]). Although they have all some similarities, they are all different in structure and vocabularies with some specific components. In the objective to provide a framework for geospatial Web services, ISO/TC 211 has worked on the preparation of a framework [67] that subsumes the mandatory elements and some optional element from these frameworks with the additional consideration from ISO 19115-1 [81] and ISO 19119 [82]. As illustrated in Fig. 17.18, the proposed framework includes GeoWeb service identification, description, capability, and metadata. The framework is also available in OWL, so that individual ser-

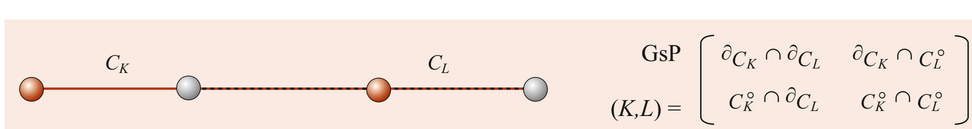


Fig. 17.17 *K* and *L* context intersection

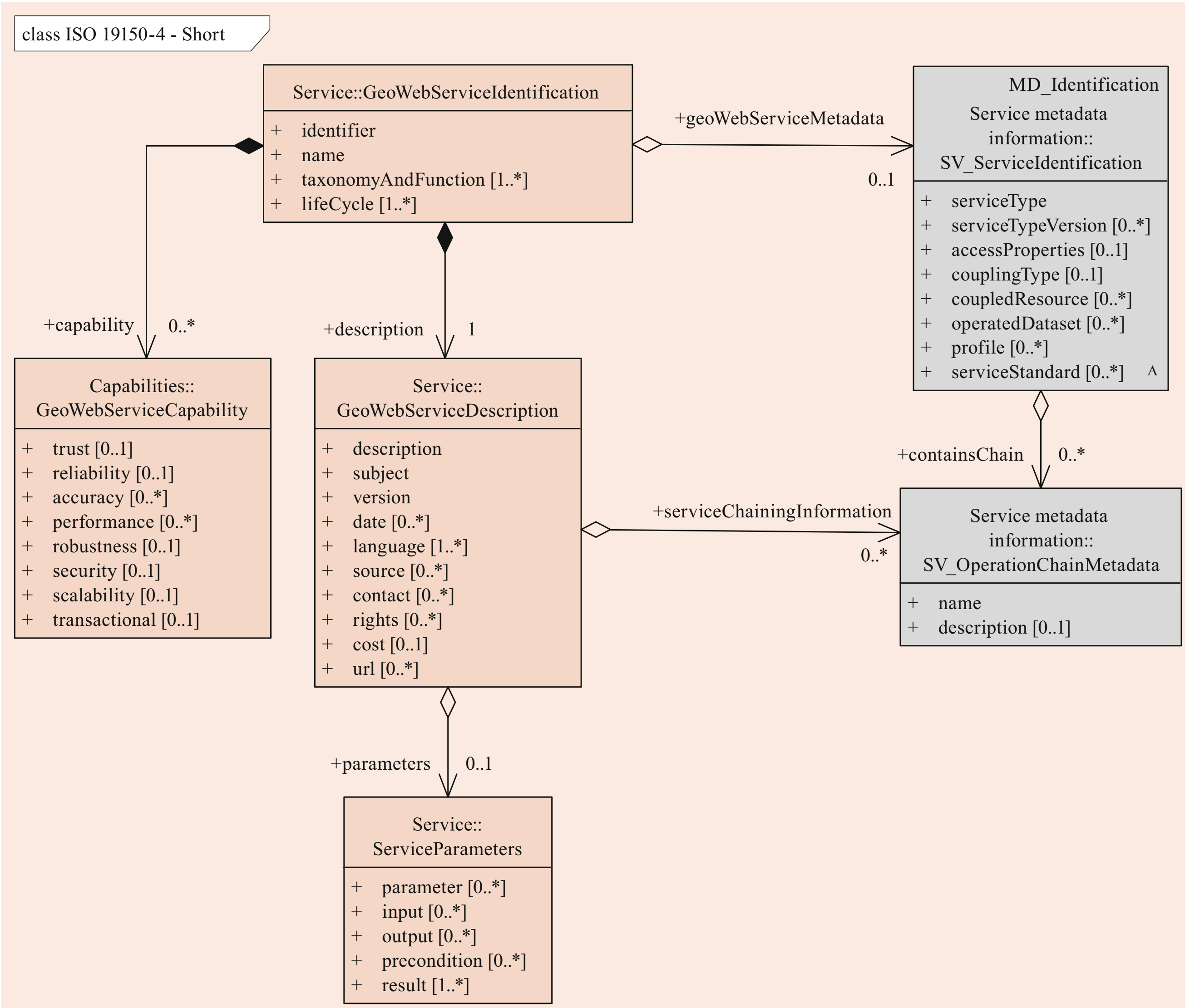


Fig. 17.18 Proposed ISO/TC 211 service ontology framework

vices can be described in compliance to the framework and be published on the Semantic Web.

17.8.2 Embedding Geographic Information in the Semantic Web

Following the idea of the Semantic Web, the Geospatial Semantic Web was tabled in 2002 [83, 84]. Its goal was to extend the notion of the Semantic Web to address the issues of, and to enhance the semantic interoperability of, geographic information over the Web. A number of challenges were then identified:

- Ontologies of spatial concepts used across disciplines (e.g., point, curve, polygon, etc.)
- Geospatial relations ontology (e.g., disjoint, touch, inside, etc.)
- Geographic feature ontology (e.g., land cover classification, road network features, hydrographic network, etc.)
- Ontology management: designing, developing, storing, registering, discovering, browsing, maintaining, and querying
- Canonical form for geospatial data queries (e.g., SPARQL for geographic information)
- Matching concepts to ontologies (e.g., geosemantic proximity assessment)
- Ontology integration.

Since that time, research and development undertaken by the research community, standardization bodies (namely ISO/TC 211 and OGC), and the geographic information industry has to a large extent already provided the anticipated results. Through international standardization activities, ISO/TC 211 has defined a comprehensive ontology of geospatial concepts that are application independent. This ontology is a foundation for describing geographic information and, as introduced in Sect. 17.7, includes concepts for the description of geometry, topology, temporal information, spatial reference systems, features, characteristics, behaviors, relations, quality, metadata, services (positioning, portrayal, location-based, etc.), imagery, gridded data, coverage, sensors, moving features, etc. Additionally, service interfaces have been elaborated to discover, access, and obtain geographic information, including the Web Map Server (WMS), Web Feature Service (WFS), filter encoding, Web Coverage Service (WCS), Catalogue Services for the Web (CSW), Web Processing Service (WPS), etc. These service interfaces define the adopted form of Web queries specifically for geographic information. The research communities in ontology, semantic interoperability, and geographic information have devoted an important effort to address the issues of concept matching and ontology integration through the study of similarity between concepts, proximity, and matching distance [14, 75, 77, 85–87]. We can foresee that the result of this research will be integrated as part of services and applications to give increased reasoning capabilities and interaction between geographic data coming from multiple sources on the Web. More recently, work on the definition of geographic feature ontologies has begun. The development of such ontologies is probably the most difficult issue to the requirements and the expertise are multifold. The development of ontology is address since based on consensual agreement. In the international geographic information arena, the development of ontologies was first initiated for domain ontology, where experts representing countries and international organizations are collaborating together. This is the case for the development of the Land Cover Classification System [88, 89], the Land Administration Domain Model [90, 91], and we can foresee that additional topics will be to be covered for the United Nations Sustainable Development Goal. Accordingly, countries, organizations, and industries will be able to undertake the derivation of more specific ontologies to address specific requirements but in compliance with upper-level ontologies. As one can recognize, the geospatial semantic agenda is clearly underway.

Furthermore, the geospatial Semantic Web brings important additional capabilities to geographic information interoperability, making geographic information available to a wider community of users. The Geospatial Semantic Web allows interoperability across domains. Users of various disciplines have greater possibilities of integrating geographic information within their specific application. This includes maps as part of tourism information, satellite images for agricultural analysis, geographic location for car services, etc. The Geospatial Semantic Web will add geographic information and spatial and temporal analysis capabilities to the Semantic Web to enhance machine reasoning and inference; for example, addresses can then be matched with location, and further, location with accidents. By enabling additional reasoning and inference, the Geospatial Semantic Web not only increases the information available on the Web but also provides new knowledge dynamically. If your preferred outdoor store has moved from one place to another, and if its location is automatically updated on its Web information, then all Web resources linked to it are aware of this and provide the new route or distance to get to your store. Geographic information now becomes an online resource that provides access to information and knowledge as opposed to offline access. It interrelates similar and/or different concepts, for example, with the use of different keywords for similar concepts in metadata. It also allows the association of similar and/or different concepts between domains. Consequently, standards in geographic information and standardized geographic information will become exposed to other communities that are not aware of the spatial domain. Hence, the Geospatial Semantic Web is more than simply enhancing the interoperability of geographic information but also allows interoperability of information across all Web data sources integrating both geographic and non-geographic information. As a result, geographic information will have higher penetration on the Web.

A number of issues now become of particular interest for the Geospatial Semantic Web. During the last few decades, the main focus has been on sharing geographic data from which users derive the information they require for their specific needs. We can certainly foresee a future oriented towards knowledge where Web data will be better structured, interrelated, and organized in a way to provide users with accurate and detailed answers to their queries directly, as opposed to simply obtaining data. The Web, or more precisely, application and services on the Web will have the capability to reason and integrate specific data and answer users' requests similarly as described in the framework for geosemantic interoperability in Fig. 17.7.

This will require access to ontologies, of course. However, ontologies will need better structures. OWL and similar languages provide a syntax to develop ontologies, but the same set of concepts can be represented differently using the same syntax. Ontologies, at least within a unique domain, will benefit when described uniformly. That way, it will be easier for applications and services to use ontologies together and benefit from the synergy this creates. Rules or best practices will then be required for ontology development.

Developing ontologies is not an aim in itself. Associating and integrating ontologies with data make spatial, temporal, and semantic reasoning and inference possible, which is an essential functionality in the Geospatial Semantic Web. As such, ontologies are part of the data lifecycle and must be developed and made available with data and data product specifications.

Additionally, the dynamic derivation of spatial, temporal, and semantic relationships between concepts and feature instances is feasible through the implementation of software agents. Software agents can relate concepts and feature instances to derive their similarity. However, this requires the definition of spatial, temporal, and semantic relations in Semantic Web languages (e.g., RDF, RDF-S, and OWL) and reasoning functions that will be integrated in software agents. ISO/TC 211 UML models translated in OWL constitute a foundation for the development of ontologies in geographic information, which would contribute to reasoning on the Semantic Web.

Another issue that is of great importance for the Geospatial Semantic Web is the development of domain ontologies describing geographic phenomena (e.g., road network, hydrologic network, relief, forest, land, geology, etc.). This is a pragmatic orientation that will enhance semantic interoperability on the Web and support ontology mapping between domains.

The Geospatial Semantic Web is also composed of geospatial Web services. Geospatial Web service ontologies will then be required to discover and access geospatial Web services. ISO 19119 sets the basis for developing geospatial services and for the documentation of these services through service metadata. This is a foundation for the definition of geospatial service ontology.

17.9 Conclusion

Geosemantic interoperability has been an area of research and development for the last two decades. The research community along with standardization bodies have contributed significantly to its development. Progress on the World Wide Web was impressive during the same period.

This chapter provides an account of semantic interoperability in the realm of geographic information. Semantics has been depicted as a triad (semiotic triangle) involving the *referent* (the phenomenon), the *signifier* (the symbols), and the *signified* (the thoughts). Cognition is significantly related to semantics, where *perceptual symbols* stored in human memory constitute the set of concepts of human beings. Computer science borrowed the notion of ontology from philosophy for the description of concepts with respect to the development of semantic interoperability. Semantic interoperability can be compared to an efficient bidirectional communication process where a user agent and a provider agent interact through queries and responses, understanding each other because of their knowledge and reasoning capabilities. Spatial data infrastructures have been developed throughout the world to implement interoperability of geographic information. The next step is to move forward in implementing geosemantic interoperability. Standards are essential in the development of geosemantic interoperability. ISO/TC 211 has done important work in this direction, but more work on ontologies will be required. The Semantic Web along with the software industry in semantics have advanced the realization of semantic interoperability. The Geospatial Semantic Web will benefit from this. Additionally, the Geospatial Semantic Web will contribute to the Semantic Web, bringing a spatial foundation and spatial ontologies supporting better interoperability across domains.

Now, most development is oriented in the context of the Web, including development for the interoperability of geographic information. The recent introduction of the Semantic Web provides another perspective for addressing geosemantic interoperability. It brings new opportunities for the interoperability of geographic information such as automatic interpretation, reasoning, and inference. The establishment of ontologies in Semantic Web languages is now required to support the Geospatial Semantic Web. Standardization bodies such as ISO/TC 211 need to play a significant role in this matter.

Interoperability of geographic information is progressing in the era of geosemantic interoperability.

References

1. Wikimapia: http://wikimapia.org (last accessed 31 Dec 2017)
2. Google maps: http://maps.google.com (last accessed 31 Dec 2017)
3. OpenStreetMap: http://www.openstreetmap.org/ (last accessed 31 Dec 2017)
4. Sycara, K., Paolucci, M., Srinivasan, N.: Combining Services and Semantics on the Web (2005). http://www.w3.org/2005/04/FSWS/Submissions/31/positionPaper.pdf
5. Merriam-Webster: Definition of Semantics. https://www.merriam-webster.com, Accessed 31 Dec 2017
6. Eco, U.: Le signe. Éditions Labor, Brussels (1988)
7. Frege, G.: Sens et dénotation. In: Imbert, C. (ed.) Écrits Logiques et Philosophiques, pp. 102–126. Seuil, Paris (1971)
8. Ogden, C.K., Richards, I.A.: The Meaning of Meaning. Harcourt, New York (1923)
9. Peirce, C.S., Deledalle, G.: Écrits sur le signe. Éditions du Seuil, Paris (1978)
10. Cherry, C.: On Human Communication: A Review, a Survey, and a Criticism. MIT Press, Cambridge (1978)
11. Barsalou, L.W.: Perceptual symbol systems. Behav. Brain Sci. **22**, 577–660 (1999)
12. Pylyshyn, Z.W.: The imagery debate: Analogue media versus tacit knowledge. Psychol. Rev. **88**(1), 16–45 (1981)
13. Pylyshyn, Z.W.: Mental imagery: In search of a theory. Behav. Brain Sci. **25**(2), 157–182 (2002)
14. Laakso, A., Cottrell, G.: Content and cluster analysis: Assessing representational similarity in neural systems. Philos. Psychol. **13**, 47–76 (2000)

15. Lakoff, G.: Women, Fire, and Dangerous Things – What Categories Reveal About the Mind. University of Chicago Press, Chicago (1987)
16. Peuquet, D., Smith, B., Brogaard, B.: The Ontology of Fields. NCGIA, Santa Barbara (1998)
17. Smith, B., Mark, D.M.: Geographical categories: An ontological investigation. Int. J. Geogr. Inf. Sci. **15**, 591–612 (2001)
18. Lehmann, F.: Semantic networks. Comput. Math. Appl. **23**, 1–50 (1992)
19. Gruber, T.R.: A Translation Approach to Portable Ontology Specification KSL 92-71. Knowledge Systems Laboratory, Stanford (1993)
20. Guarino, N.: Formal ontology and information systems. In: Proc. Form. Ontol. Inf. Syst., Trento, pp. 3–15. (1998)
21. Brodeur, J., Bédard, Y., Edwards, G., Moulin, B.: Revisiting the concept of geospatial data interoperability within the scope of a human communication process. Transactions GIS **7**, 243–265 (2003)
22. Kashyap, V., Sheth, A.: Semantic and schematic similarities between database objects: A context-based approach. VLDB Journal **5**, 276–304 (1996)
23. Mackay, D.S.: Semantic integration of environmental models for application to global information systems and decision making. ACM SIGMOD Rec. **28**, 13–19 (1999)
24. Daconta, M.C., Obrst, L.J., Smith, K.T.: The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management. Wiley, New York (2003)
25. Bergamaschi, S., Castano, S., Vincini, M.: Semantic integration of semistructured and structured data sources. ACM SIGMOD Rec. **28**, 54–60 (1999)
26. Kahng, J., McLeod, D.: Dynamic classificational ontologies: Mediation of information sharing in cooperative federated database systems, context and ontologies. In: Papazoglou, M.P., Schlageter, G. (eds.) Cooperative Information Systems – Trends and Directions, pp. 179–203. Academic Press, San Diego (1998)
27. Kashyap, V., Sheth, A.: Semantic heterogeneity in global information systems: The role of metadata, context and ontologies. In: Papazoglou, M.P., Schlageter, G. (eds.) Cooperative Information Systems – Trends and Directions, pp. 139–178. Academic Press, San Diego (1998)
28. Sheth, A.: Changing focus on interoperability in information systems: From systems, syntax, structure to semantics. In: Goodchild, M., Egenhofer, M., Fegeas, R., Kottman, C. (eds.) Interoperating Geographic Information Systems, pp. 5–29. Kluwer Academic, Boston (1999)
29. Smith, B., Mark, D.: Ontology with human subjects testing: An empirical investigation of geographic categories. Am. J. Econ. Sociol. **58**, 245–272 (1999)
30. Fowler, J., Perry, B., Nodine, M., Bargmeyer, B.: Agent-based semantic interoperability in InfoSleuth. ACM SIGMOD Rec. **28**, 60–67 (1999)
31. Smith, H., Poulter, K.: Shared ontology in XML-based trading architecture. Communications ACM **42**, 110–111 (1999)
32. Smith, K., Obrst, L.: Unpacking the semantics of source and usage to perform semantic reconciliation in large-scale information systems. ACM SIGMOD Rec. **28**, 26–31 (1999)
33. Sycara, K., Klusch, M., Widoff, S., Lu, J.: Dynamic service matchmaking among agents in open information environments. ACM SIGMOD Rec. **28**, 47–53 (1999)
34. Princeton University: About WordNet. Princeton Univ, Princeton (2010). http://wordnet.princeton.edu
35. Geraci, A.: IEEE Standard Computer Dictionary, A Compilation of IEEE Standard Computer Glossaries 610. IEEE, New York (1991)
36. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
37. Shannon, C.E.: A mathematical theory of communication (Concluded from July 1948 issue). Bell Syst. Tech. J. **27**(4), 623–656 (1948)
38. Weiner, N.: The Human Use of Human Beings: Cybernetics and Society. Houghton Mifflin, San Diego (1950)
39. Schramm, W.: How communication works. In: DeVito, J.A. (ed.) Communication: Concepts and Processes, pp. 12–21. Prentice Hall, Englewood Cliffs (1971)
40. Bishr, Y.: Semantics Aspects of Interoperable GIS. Dissertation. ITC Publication, Enschede (1997)
41. Ouksel, A.M., Sheth, A.: Semantic interoperability in global information systems: A brief introduction to the research area and the special section. ACM SIGMOD Rec. **28**, 5–12 (1999)
42. Bishr, Y.: Overcoming the semantic and other barriers to GIS interoperability. Int. J. Geogr. Inf. Sci. **12**, 299–314 (1998)
43. Bédard, Y., Proulx, M.-J.: Perceptory. Université Laval, Québec (2002). http://sirs.scg.ulaval.ca/Perceptory (last accessed 31 Dec 2017)
44. Sowa, J.F.: Conceptual Structures: Information Processing in Mind Machine. Addision-Wesley, Reading (1984)
45. Bédard, Y.: A Study of the Nature of Data Using a Communication-Based Conceptual Framework of Land Information. Dissertation. Univ. Maine, Orono (1986)
46. Natural Resources Canada: Standards user's guide for geographic information. Canadian Geospatial Data Infrastructure, Information Product 46e (2016). https://doi.org/10.4095/297543, Accessed 31 Dec 2017
47. Global Spatial Data Infrastructure Association: Spatial Data Infrastructure Cookbook 2012 Update (2012). http://gsdiassociation.org/images/publications/cookbooks/SDI_Cookbook_from_Wiki_2012_update.pdf, Accessed 31 Dec 2017
48. Hjelmager, J., Moellering, H., Cooper, A., Delgado, T., Rajabifard, A., Rapant, P., Danko, D., Huet, M., Laurent, D., Aalders, H., Iwaniak, A., Abad, P., Düren, U., Martynenko, A.: An initial formal model for spatial data infrastructures. Int. J. Geogr. Inf. Sci. **22**, 1295–1309 (2008)
49. ISO/IEC 10746:1998: Information Technology – Open Distributed Processing – Reference Model: Overview. International Organization for Standardization, Geneva (1998)
50. US DOD: Joint Publication 3-13. Joint Chiefs of Staff, Washington (2012). http://www.dtic.mil/doctrine/new_pubs/jp3_13.pdf (last accessed 31 Dec 2017)
51. ISO 19109:2015: Geographic Information – Rules for Application Schema. International Organization for Standardization, Geneva (2015)
52. ISO 19110:2016: Geographic Information – Methodology for Feature Cataloguing. International Organization for Standardization, Geneva (2016)
53. Open Geospatial Consortium Inc.: OpenGIS Web Map Server Implementation Specification, OGC 06-042, Version: 1.3.0, 2006. OGC, Wayland (2006). http://www.opengeospatial.org/standards/wms (last accessed 31 Dec 2017)
54. Open Geospatial Consortium Inc.: OGC® Web Feature Service 2.0 Interface Standard – With Corrigendum, OGC 09-025r2, Version: 2.0.2, 2014. OGC, Wayland (2009). http://www.opengeospatial.org/standards/wfs (last accessed 31 Dec 2017)
55. Open Geospatial Consortium Inc.: WCS 2.0 Interface Standard – Core, OGC 09-110r4, Version: 2.0.1, 2012. OGC, Wayland (2009). http://www.opengeospatial.org/standards/wcs (last accessed 31 Dec 2017)
56. Open Geospatial Consortium Inc.: OGC® WPS 2.0 Interface Standard, OGC 14-065, Version: 2.0, 2015. OGC, Wayland (2015). http://www.opengeospatial.org/standards/wms (last accessed 31 Dec 2017)

57. ISO/TC 211: Geographic Information/Geomatics. International Organization for Standardization, Geneva (2017). https://committee.iso.org/home/tc211
58. ISO 19101-1:2014: Geographic Information – Reference Model. International Organization for Standardization, Geneva (2014)
59. OMG: OMG Unified Modeling Language (OMG UML) (Version 2.5). Object Management Group, Needham (2015)
60. World Wide Web Consortium: OWL 2 Web Ontology Language Primer, 2nd edn. W3C, Boston (2012). W3C Recommendation 11 Dec 2012, http://www.w3.org/TR/owl-primer/ (last accessed 31 Dec 2017)
61. ISO/TS 19150-1:2012: Geographic Information – Ontology – Part 1: Framework. International Organization for Standardization, Geneva (2017)
62. ISO/TS 19150-2:2015: Geographic Information – Ontology – Part 2: Rules for Developing Ontologies in the Web Ontology Language (OWL). International Organization for Standardization, Geneva (2015)
63. ISO 19107:2003: Geographic Information – Spatial Schema. International Organization for Standardization, Geneva (2003)
64. ISO 19108:2002: Geographic Information – Temporal Schema. International Organization for Standardization, Geneva (2002)
65. ISO 19125-1:2004: Geographic Information – Simple Feature Access – Part 1: Common Architecture. International Organization for Standardization, Geneva (2004)
66. ISO 19141:2008: Geographic Information – Schema for Moving Features. International Organization for Standardization, Geneva (2008)
67. ISO/CD 19150-4:2017: Geographic Information – Ontology – Part 4: Service Ontology. International Organization for Standardization, Geneva (2015)
68. ISO/TC 211: Geographic Information/Geomatics: ISO/TC 211 Resources. International Organization for Standardization, Geneva (2017). https://committee.iso.org/sites/tc211/home/re.html (last accessed 31 Dec 2017)
69. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Sci. Am. **284**, 34–43 (2001)
70. Berners-Lee, T.: Artificial intelligence and the semantic web. In: Proc. Conf. Artif. Intell. AAAI-06 (2006)
71. CanVec: Topographic Data of Canada. http://open.canada.ca/data/en/dataset/8ba2aa2a-7bb9-4448-b4d7-f164409fe056
72. Berners-Lee, T.: Uniform Resource Identifier (URI), Generic Syntax. Network Working Group, IETF, Fremont (2005). http://tools.ietf.org/html/rfc3986 (last accessed 31 Dec 2017)
73. ISO/IEC 10646:2017: Information Technology – Universal Coded Character Set (UCS). International Organization for Standardization, Geneva (2017)
74. World Wide Web Consortium: W3C – World Wide Web Consortium. W3C, Boston (2017). http://www.w3.org/
75. Euzenat, J., Shvaiko, P.: Ontology Matching. Springer, Berlin, Heidelberg (2007)
76. Egenhofer, M.J., Herring, J.R.: Categorizing binary topological relations between regions, line and points in geographic databases. In: Egenhofer, M., Mark, D.M., Herring, J.R. (eds.) The 9-Intersection: Formalism and Its Use for Natural-Language Spatial Predicates, pp. 1–28. NCGIA, Santa Barbara (1994). Technical Report
77. Brodeur, J., Bédard, Y., Moulin, B.: A geosemantic proximity-based prototype for interoperability of geospatial data. Comput. Environ. Urban Syst. **29**, 669–698 (2005)
78. World Wide Web Consortium: OWL-S: Semantic Markup for Web Services. W3C, Boston (2004). W3C Member Submission 22 Nov 2004, http://www.w3.org/Submission/OWL-S (last accessed 31 Dec 2017)
79. World Wide Web Consortium: Semantic Web Services Ontology (SWSO). W3C, Boston (2005). W3C Member Submission 9 Sep 2005, http://www.w3.org/Submission/SWSF-SWSO/ (last accessed 31 Dec 2017)
80. World Wide Web Consortium: Web Service Modeling Ontology (WSMO). W3C, Boston (2005). W3C Member Submission 3 Jun 2005, http://www.w3.org/Submission/WSMO/ (last accessed 31 Dec 2017)
81. ISO 19115-1:2014: Geographic Information – Metadata – Part 1: Fundamentals. International Organization for Standardization, Geneva (2014)
82. ISO/TS 19119:2016: Geographic Information – Services. International Organization for Standardization, Geneva (2016)
83. Egenhofer, M.J.: Toward the semantic geospatial web. In: Proc. 10th ACM Int. Symp. Adv. Geogr. Inf. Syst, pp. 1–4. ACM, New York (2002)
84. Fonseca, F., Sheth, A.: The Geospatial Semantic Web. UCGIS, Leesburg (2002). White Paper
85. Tversky, A.: Features of similarity. Psychol. Rev. **84**(4), 327–352 (1977)
86. Rodriguez, M.A., Egenhofer, M.J., Rugg, R.D.: Assessing semantic similarities among geospatial feature class definition. Lect. Notes Comput. Sci. **1580**, 189–202 (1999)
87. Formica, A., Pourabbas, E.: Content based similarity of geographic classes organized as partition hierarchies. Knowl. Inf. Syst. **20**(2), 221–241 (2009)
88. Di Gregorio, A.: Land Cover Classification System – Classification Concepts and User Manual. FAO, Rome (2005). software version 2
89. ISO/DIS 19144-2: Geographic Information – Classification Systems – Part 2: Land Cover Meta Language (LCML). International Organization for Standardization, Geneva (2010)
90. Hespanha, J., Van Bennekom-Minnema, J., Van Oosterom, P., Lemmen, C.: The model driven architecture approach applied to the land administration domain model version 1.1 – with focus on constraints specified in the object constraint language. In: FIG Working Week, Stockholm (2008)
91. ISO 19152:2012: Geographic Information – Land Administration Domain Model (LADM). International Organization for Standardization, Geneva (2012)

Jean Brodeur Jean Brodeur received a PhD in Geomatics Science in 2004. He has worked for Natural Resources Canada (1985–2015) and been adjunct at Université Laval (2005–2016) and member of the Centre de recherche en Géomatique. In 2015, he founded GéoSémantic Research and provides consultancy in conceptual modeling and geographic databases, geographic information systems, standardization, interoperability, metadata, and the Semantic Web.

18 Registration of Geospatial Information Elements

C. Douglas O'Brien and Roger Lott

Contents

C. D. O'Brien (✉)
IDON Technologies, Inc.
Ottawa, Ontario, Canada
e-mail: cdobrien@idontech.ca

R. Lott
Ilkley, West Yorkshire, UK
e-mail: rogerlott@btinternet.com

Abstract

Geospatial information elements are the fundamental constituents of geographic information. A dataset of geospatial information includes the geographic features, their attributes, and the relationships between these features together with the associated metadata and spatial elements. Most of the suite of geospatial standards define templates for describing these features, attributes, relations, and metadata. The actual feature, attribute, and metadata information element definitions are application dependent; for example, a road may be described differently on a transportation logistic map than on a cadastral map, and more detailed attributes about trees may exist on a forestry land cover map than on a political jurisdiction map. Registers may be used to hold these definitions so that they may be used in common in similar applications. In fact, the largest part of the standardization of geographic information, with respect to the management of detail, is the standardization of feature and attribute dictionaries and common code lists that may be used to support attributes and metadata. This chapter describes how ISO standards 19135-1 and 19135-2 provide for the management of registers of geographic information elements, including the complex relationships between multiple registers. Registers allow for the management of diversity.

Keywords

Georegister · Feature Register · Coordinate System Register

18.1 Background

With the expansion of digital geospatial information into more application areas, and especially the development of geospatial information services, it is important to have common information elements on which to base data products

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_18

and services. The old approach of implicitly defining information components within an exchange standard has become unworkable due to the need for a more responsive means of managing the information elements. The more modern approach is to define the information independently of the exchange format, maintaining a separation between the *information content* and the *carrier* of that content. This allows for multiple different exchange or storage media to be used with the same information.

There are two major aspects to the *information content*: the first is the definition of the information content elements, and the second is the relationship between the elements. The definition of the meaning of the set of information elements is a list, which is best addressed in a register. The structure of each element and the relationship between elements can be managed in a schema.

18.2 Requirements

Information systems have become much more interconnected over the past 30 years. In the early days it was sufficient to be able to exchange information that would assist production agencies in their own development of their map products. Gradually this evolved into the establishment of common information products that could be distributed directly to end-users. In the past decade, this has further evolved into the establishment of information services based on selections of information from multiple sources. These modern information services require a high level of commonality within the data available from different mapping and other agencies. However, at the same time, the rate of evolution of technology has increased, so that it is much more difficult to create a high level of commonality. This is especially true for *bridge* services that integrate material from many sources, such as a web feature service. In a web feature service (WFS), data may be integrated from many different information sources that may contain similar but not identical features. For example, the shoreline feature on a topographic map is elevation contour 0 related to the elevation datum of the topographic map. The shoreline on a hydrographic chart is based on a different hydrographic datum such as the mean sea level, the average level of the constantly moving sea. The shoreline on a chart is *not the same* as that on an adjacent topographic map, and the two cannot simply be integrated. A mechanism is needed to bridge between the definitions of these two information elements if they are to be used together.

When we integrate information from many sources we must accommodate:

- *Different, independent information domains.* There cannot be one comprehensive standardized information domain, because every national and international organization has different interests. Some things are more important to certain players than to others. Since not all participants have the same motives, or schedules, there will never be agreement on a comprehensive solution. A federated solution that permits diversity is required.
- *Different rates of development.* The rate of development depends upon the level of effort expended, which depends upon budgets and other resources expended. Because different nations or organizations have different needs, the rate of development will vary considerably. This means that integrated systems will need to accommodate older systems and newer systems and legacy and new data at the same time. This establishes both a backward and a forward compatibility requirement.
- *Broad applicability.* The concept of what is geospatial information is getting ever broader. This has strong implications for agreements already made in existing standards. As new information domains are integrated into systems, the meaning of existing information components may change because the context becomes broader.
- *Wide availability.* A register must be open and freely available to users and external organizations that participate in the federated approach.
- *Different languages.* A register accommodates cultural and linguistic adaptability by allowing different definitions in different parallel registers to be linked by a common item identifier (ID); for example, a system built in Germany and one in Canada might both include a geographic feature called a *Road*. The German system might have a German-language definition, and the Canadian system have both an English- and French-language definition of the concept. If the definition is registered, then the common item ID can bridge between the three languages in use in this example.

The approach to handling these requirements is to establish a federated system of registers that allow for independence of the different information providers, as well as appropriate cross-referencing so that maintenance can be managed.

18.3 Concept of a Register

A register is a managed list that contains information elements. A schema is a set of relations between information elements from that list. Registers may be used to manage many different types of information. There are a few common elements that are needed to ensure that a register is well maintained and supports backward compatibility and external referencing. Otherwise a register may contain any attributes required to define an information element. A general standard for managing metadata registries is defined in ISO/IEC standard 11179 [1].

ISO standard 19135-1 [2] specifies procedures for the registration of items of geographic information. The use of ISO 19135-1 as the base for registration provides a number of benefits including ease of updating and the management of temporal change, a common interface, support for cultural and linguistic adaptability, as well as the commonality inherent in the use of a standardized approach.

ISO 19135-1 permits the establishment of hierarchical registers. A hierarchical register is a structured set of registers composed of a principal register and a set of subregisters. Hierarchical registers may be of many levels. Each subregister is itself a register, which may also contain subregisters. A specification is required to define the contents of a register and the relation to any subregisters.

A register never loses any information. Elements may be added to the register but are never removed. The status may change from valid to retired or superseded, but the element remains in the register. This means that old systems or data may continue to reference the register using the original item identifier and still reference the same element. This satisfies the important backward compatibility requirement. Systems and data that are compliant using elements from a register remain compliant even if the register is updated, because the older element definitions remain in the register. A pointer that indicates the new element replacing a superseded element in a register assists in providing for forward compatibility; that is, data may be converted based on following the links imbedded in the register.

In accordance with ISO 19135-1, every element in a register is marked with the date it was entered into the register. If the element has been retired or superseded, it is also marked with the date it was changed to that status. This means that a register can support truly dynamic evolution without making obsolete systems or data products that depend upon it.

A proposed element may also be entered into a register with the state *notValid*. The *dateAccepted* attribute would be entered when the item is approved, and the state is changed to *Valid*. If it is not approved, it will still remain in the register, using up an instance of *itemIdentifier*, since elements are never removed. This is beneficial because old proposals can be reviewed to ensure that new proposals are handled in the same manner.

The ISO standard 19135:2005 [3] has been revised to become ISO 19135-1:2015. This revision has not changed the concept of a register, and the revision is entirely backward compatible. The revision broadened the concepts to allow additional information types to be registered. In addition, a second part to the standard was developed, ISO 19135-2 [4], which provides an XML schema for the exchange of registered information; that is, to allow the contents of a register to be communicated, to facilitate backup/archive, and to allow integration of material from one register into another.

18.4 Register Versus Registry

The terms register and registry are, unfortunately, very similar, but they have very different meanings. ISO 19135-1 defines:

- *Register:* set of files containing identifiers assigned to items with descriptions of the associated items
- *Registry:* information system on which a register is maintained

A register represents the information content, whereas a registry is a database or other computer system that holds the content.

In normal conversation, people do not come across the detailed meanings of these two common English words, but they do hold these meanings in other contexts. For example, the Land Registry Office (a building) contains a book that contains the Register of Land Holdings (a list). To add to the confusion, some systems for implementing registers have used these terms in exactly the opposite way.

18.4.1 Versioning of Registers

The date mechanism inherent in the register mechanism defined by ISO 19135-1 means that the registration process is an ongoing one. Information can be added at any time without conflict. This mechanism is important, especially in the management of a federated set of registers.

When a nation or organization develops a service or product, it defines a service or product specification that references a register or a set of registers at a given date. This fixes the instances of register entries used in that product. This approach works, and the product remains valid even if the register changes over time.

If a nation or organization wishes to develop several services or data products, it is desirable that they all reference the set of registers at the same date. Minor differences might occur in the register content if two products or services reference registers only a few weeks apart, and this might require conversion in operation of the services or use of the data. Therefore, versioning of registers can be introduced. Versions establish arbitrary baselines that can be referenced by a number of different products or services. If a product or service requires an extension beyond a baseline version, this is of course allowed, but the use of versioning reduces unnecessarily small differences between products and/or services.

18.5 Registration Process

ISO standard 19135-1 indicates that it is "intended particularly to apply to registers established under the auspices of ISO/TC 211" but it also indicates that "any organization may choose to establish registers of items of geographic information that conform to this part of ISO 19135". Any organization can establish its own set of registers under its own authority in compliance with all the requirements set out in ISO 19135-1. These registers will be compatible with ISO/TC 211 registers and also registers developed by other organizations in compliance with the ISO standard. This is important for establishing a federated network of linked registers.

To allow for a federated network of linked registers, one should make use of the optional attributes of *authoritative reference source* and *lineage*. These attributes allow the user of a register to trace from a register item to the authority that is maintaining an externally referenced element and also separately to trace back to the lineage of the item. These two references are not always the same; for example, a particular item might be taken from the Open Geospatial Consortium (OGC), but the OGC may have taken this same item from a different ISO standard. The ISO standard is the authoritative reference. It is important to know from where the authority is derived in case that standard ever changes, but it is also important to know that the OGC is using the same element. The two links are different.

The roles that must be addressed in a registration process are given in ISO 19135-1. The following is a summary.

18.5.1 Register Owner

The register owner is the organization responsible for a register or set of registers, which as such would be responsible for managing an appeals process for submissions made to the control bodies for particular registers. For most types of registers, appeals are expected to be very rare, but the appeals process is part of the overlaying technopolitical process that governs the registration process and ensures that the register is an acceptable source of definitions for users. How users perceive the impartiality and utility of the register is determined by the way the register owner manages the register.

18.5.2 Submitting Organization

The submitting organizations for a register are those organizations that have made agreements with the register owner to be allowed to submit proposals for the inclusion of new material or changes to the register. Who is allowed to submit proposals also determines how the users perceive the impartiality and utility of the register.

18.5.3 Control Body

A control body is a group of technical experts appointed by a register owner to render a technical decision on each proposal. Such a decision may, of course, be appealed. One control body may control the content of one or several registers. The control body is also responsible at the technical level for managing the relationship with external registers. Control bodies should be organized upon the basis of required expertise. A specialized list, such as the list of geodetic codes and parameters, may need a specialized control body because it would require unique expertise, even though logically it is a subtype of metadata.

18.5.4 Register Manager

The register manager is the person responsible for the management of a register. Each register, even very stable minor code list registers, must have a register manager. One register manager and maintenance team may handle several registers.

18.5.5 Registry Manager

The registry manager is the person responsible for the day-to-day maintenance of that information system. In simple terms, the register is the content and the registry is the information system that holds the content. One registry may contain several registers. There may also exist several registries within an organization. The distinction between different registries is based on the user interface. If the user interface is different, and the registries therefore appear separate to the user, they should be considered as separate registries. Since one registry manager may manage several registries, a further distinction is not necessary.

18.5.6 Register User

The register user is the end-user of the information in one or many registers. There may be different categories of users, varying from system developers to end-users of geographic information products. Different user interfaces may be required for different categories of users. These user interfaces may range in complexity from an online service interface to an electronic download of all or a portion of the register in a form such as Extensible Markup Language (XML) file. All interfaces are not needed for all registers. For example, a web-based portrayal service might allow a web feature (mapping) service to render the portrayal for particular features with their geometry, whereas a download for a code list might be an XML-structured file.

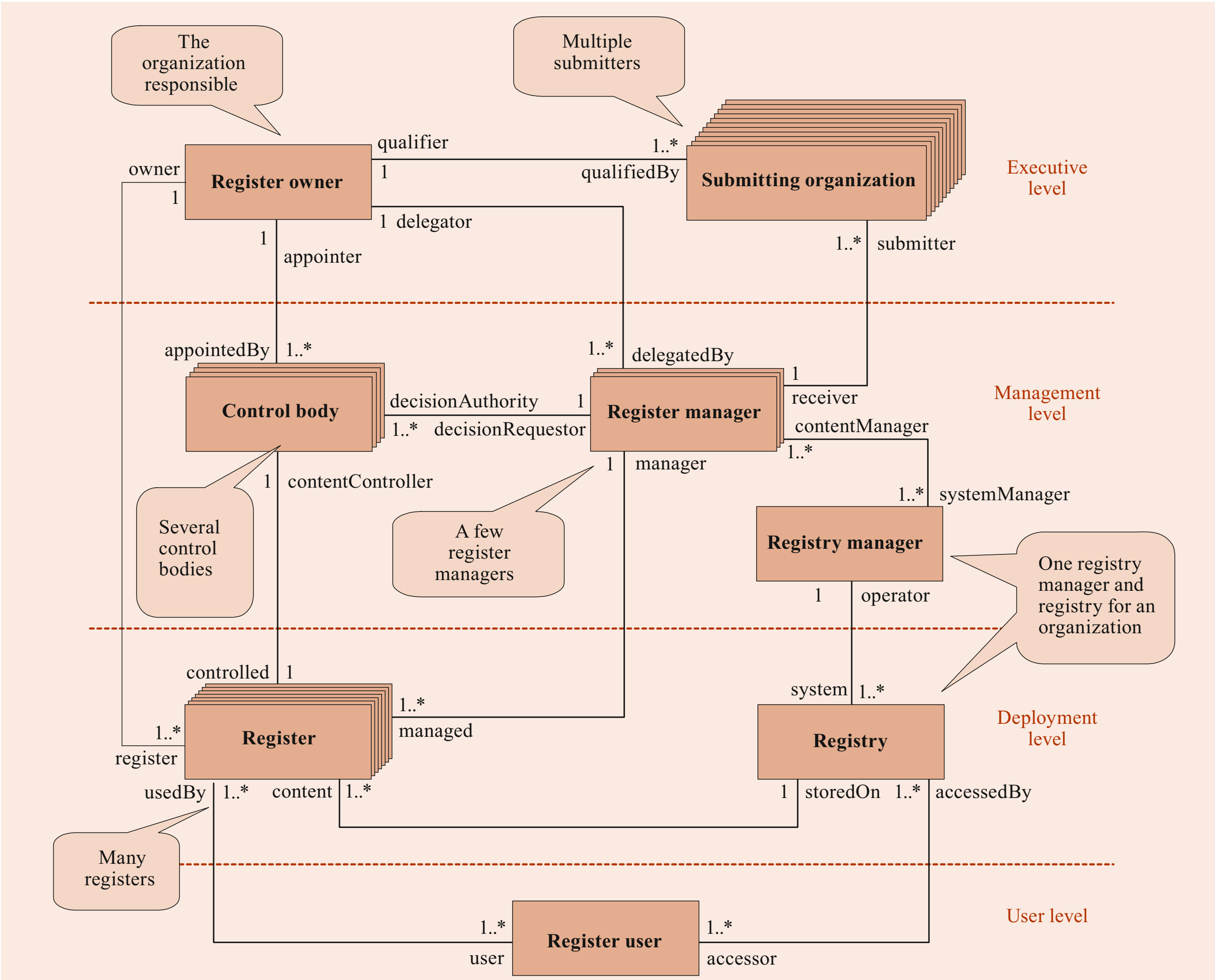


Fig. 18.1 Registration roles

The relationship between these roles is illustrated in Fig. 18.1. Note that there is only one register owner shown, but there are possibly many submitting organizations. There are a few register managers and several control bodies. A register manager and a control body make up a maintenance team for one or several registers. The make-up of a control body is primarily driven by the need to have the appropriate expertise involved.

18.6 Register Structure

18.6.1 Elements Inherited from ISO 19135-1

ISO 19135-1 specifies how registers shall be managed and the basic information that shall be included in any geographic information register. The basic register schema is shown in Fig. 18.2. In this example, three classes have been added to the reference schema defined in ISO 19135-1. These classes are subtypes of the corresponding classes in the ISO standard.

The register schema is extended to include a subtype of RE_Register for a specific-purpose register (Geo_Register), a subtype of RE_RegisteredItem corresponding to a Geo_RegisteredItem, and a subtype of RE_ItemClass corresponding to a Geo_ItemClass. Note that this extension is done in a manner parallel to that for feature items, as described in ISO 19126 [5].

18.7 Federated Registers

18.7.1 Set of Registers

An organization may require a number of different registers to address different aspects of geographic information. These

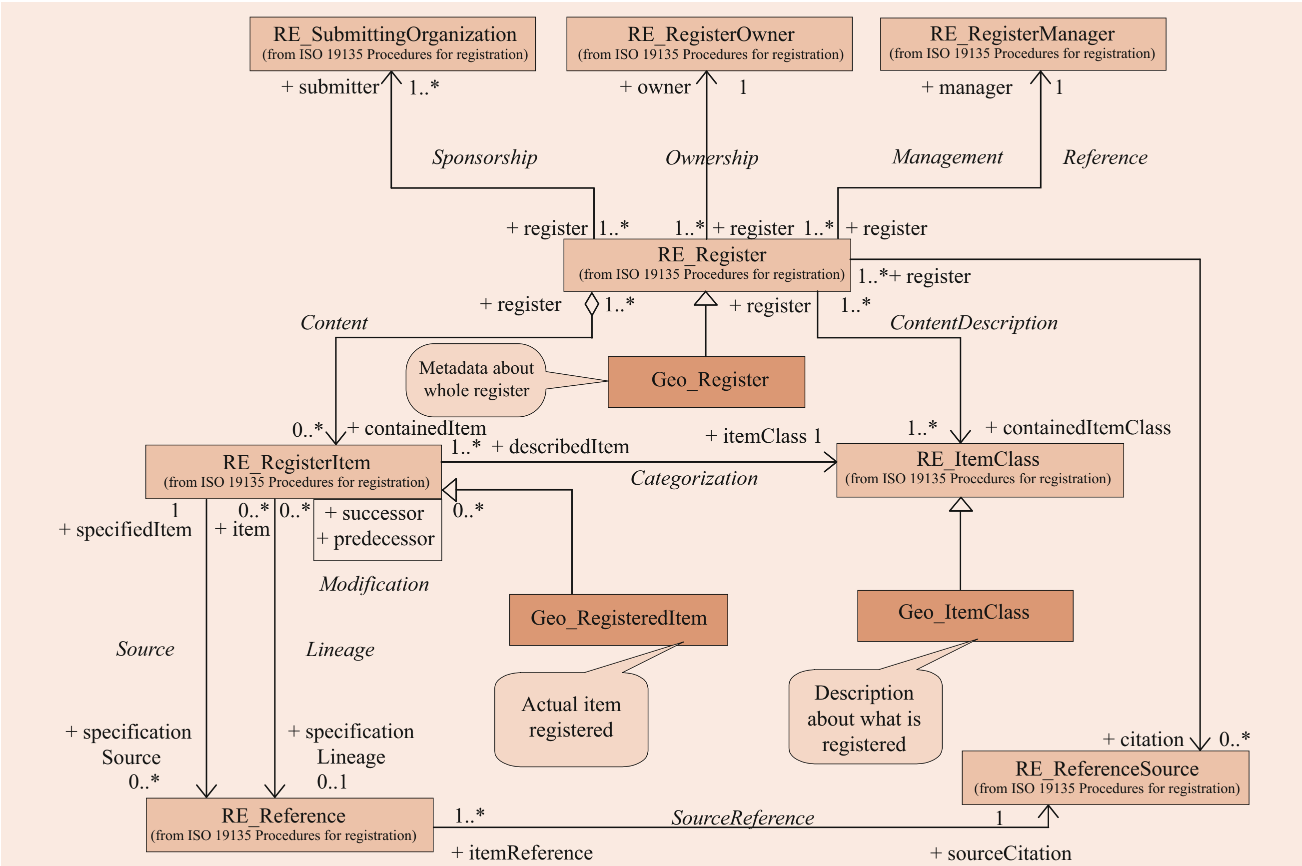


Fig. 18.2 Terminology register schema showing added classes

may vary from some simple, relatively stable code lists to large dynamic registers, such as the definition of geographic features or portrayal. All these need to be managed in the same way so that they can be used in the same information systems.

Four general classes of registers are common in geographic information systems:

- Features (real-world representation)
- Metadata (data about the data)
- Portrayal (presentation)
- Enterprise (information about the management of the standards, specifications or documentation of the data, such as common terminology).

These are all hierarchical registers; that is, there are many logical subregisters of each topic area. This is especially true of metadata, which is a catch-all, since anything that does not fit into one of the other three categories will fit into metadata. A metadata register will consist of many subregisters. Many are specialist code lists, such as the list of coordinate reference system (CRS) components (geodetic codes and parameters). Each subregister may have component subregisters; for example, there may be multiple subregisters of a CRS component register:

- Ellipsoids
- Geodetic datums and reference frames
- Vertical datums (including sounding datums)
- Coordinate transformations
- Map projections.

Within an organization there may need to be several control bodies (maintenance teams) managing these registers. One needs a separate control body for each different topic area that requires specialized expertise; for example, the CRS register would need a separate control body from other parts of metadata, because specialized expertise is required. This is the important criterium, i.e., the expertise required, and this should define what is a logical register.

There is also a requirement for managing groups of registers. Logically, these are subregisters themselves in a hierarchy. The best example comes from the portrayal area, where there will need to be subregisters of rules and subregisters of symbols. However, there will be sets of symbols that will work together on one display medium (printing versus display on a screen), and there will be sets of rules that work together with specific symbol sets to address specific application needs, such as nighttime viewing. Logically, all are registers in their own right and are subregisters of a general portrayal register. They are all sets of logical elements, and set theory allows sets of sets and even overlapping sets.

18.7.2 Reference to External Registers

The registration approach is being used in several different organizations addressing geospatial information. To be able to interoperate there is a need to harmonize these registers. The simple solution to having interoperability is to have a single, agreed register used by several organizations. This approach cannot work in general because different organizations have different interests and timelines. The relationship between these organizations will need to be managed.

Although it may sound complex, the solution is a federated approach. Different organizations will have different sets of registers. This pertains to the logical registers (the actual lists of data) not registries (the database implementations). The structure of the logical registers and the registries are separate problems, although the implementation must fit the logical structure in the end.

Registers in different organizations will need to be harmonized but cannot be identical to any particular organization's registers, because they will always be out of step and address somewhat different needs; for example, an organization can agree to accept the International Hydrographic Organization (IHO) definitions for sounding datums and for certain features, such as buoys, and make it the authoritative reference. If the IHO changes its definition, it should inform its liaison partners, who *may* decide to accept the change and follow; that is, an organization might make an authoritative reference to a particular IHO definition of a buoy. If the IHO changes its definition, data produced by the organization will still reference the old version in the IHO register. Compatibility is maintained. Products still reference the organization's registered definition and inherit the IHO definition through authoritative referencing. Existing IHO products would still reference this old definition as well. A revision of the organization's register in due time may (should) be updated to match the new IHO definition. Old products will still reference the old definitions; however, new products in both organizations will follow the change and remain harmonized. If there is a really good reason, the two organizations could decide to diverge and no longer have the IHO be the authoritative reference on a particular item.

Organizations can develop their own specific local registers as well. These may be mirror copies of the base registers maintained by other international organizations where each item makes a logical authoritative reference to the higher-level register. The reason for the local registers is to allow certain organizations to make their own additions as they feel necessary and to get out of step with the more international register. Getting out of step will happen and must be managed. It cannot just be ignored. In fact, it is beneficial. Organizations need to be able to respond easily and quickly to local needs. Over time, they can propose their local additions or modifications to the master register that they reference. If the change is accepted, that is fine; if not, the change can remain local without causing incompatibilities for others.

Registries are implementations. There will not be one overall master registry that gathers all registers. This might be convenient for users, but it is unlikely to happen. Since one database can point to another, there is really no reason to have only one implementation. A distributed implementation is not only possible but desirable. Different organizations can have their own implementations, managed and paid for by their own participants. Organizations can also mirror (copy) other organization's registers. This is the preferred approach to support secure and redundant networks, where there will be a complete copy of the required registries on an organization's local (and sometimes private) network. It can even be allowed for these mirrors to get somewhat out of step, as long as the authoritative referencing mechanism works at the logical (register) level; that is, maintenance can be periodic.

As long as we use authoritative referencing appropriately (as separate from source lineage historical referencing), it is straightforward to manage federated registers. An authoritative reference is a reference to a body that is responsible for managing a particular item in a register.

18.7.3 Register Maintenance and Access

To reduce the possibility of database corruption (possibly from hackers or maintenance problems) it is desirable that maintenance be done offline and uploaded to the registry. That is, the master database should be on a separate (well backed-up) offline machine, and the entire database (or portions thereof) should be uploaded to the registry.

There is also a need to be able to interface with the registry (database) to be able to extract the contents (register contents). There are several use cases. One is to just search and examine individual items (browse). Another is to download selected parts or the whole database of register elements, requiring an exchange format such as XML. A third is the use of the register within a service, such as a portrayal service. The purpose of the new part to the ISO Geographic information registration standard 19135-2 is to provide an XML schema to facilitate electronic download or exchange of the content of a register.

18.8 Implementation of Registers

18.8.1 Platform Independence

A register is just a managed list, and the procedures defined in the ISO 19135-1 standard are management procedures independent from how a register is implemented. A very simple register could be implemented *on paper* as a simple text file which could be edited manually (in compliance with ISO 19135-1). This may be sufficient if the list is relatively short and stable. An easier to use implementation might be an XML file that could be more easily read by computer systems. Large registers or registers that are frequently updated or accessed online require more sophisticated tools. This could be a simple database. If the register is complex or needs to be interfaced to an online web service, sophisticated tools exist, such as the e-business registry information model (ebRIM).

ebRIM is another ISO standard, ISO/TS 15000-3: 2004 [6], based on the work of the Organization for the Advancement of Structured Information Standards (OASIS) [7]. The ebRIM registry information model provides a way to describe the registry content. The OASIS standard ebXML provides a standard input/output XML format including registry services and protocols (ebRS) for developing business registry applications. ebRIM defines a number of registry objects and a method of specifying the relationship between these objects. It can be used to implement an ISO 19135-1-compliant registry; however, it is not the only implementation technique. It may not be cost-effective for small registers.

18.8.2 Geospatial Web Services

The Open Geospatial Consortium (OGC) has selected the OASIS ebRIM standard as the preferred implementation model for implementation of the OpenGIS Catalogue Service for the Web (CSW). The CSW allows users to search for geospatial data based on metadata defined in accordance with the ISO 19115-1 [8] and ISO 19115-2 [9] metadata standards. ebRIM-based registry technology allows the development of federated registers that support CSW. ebRIM-based registers will also support other OGC register requirements, such as portrayal symbol sets, coordinate reference systems, application profile, and schema registers and other registers used within the OGC specifications for web services.

18.8.3 Implications of ebRIM

The registry information model (ebRIM) is a metamodel for data within a registry. It describes information elements and their relations. The metamodel is very flexible. The importance of the metamodel lies in its specification of a common model for understanding, sharing, and reusing the contents of a registry implementation. Unified Modeling Language (UML) [10] could be used as the high-level metamodel for an organization's registers, but there is some benefit in using the ebRIM metamodel, since this model specifies the relations in more detail.

18.9 Example Registers

The discussion of federated registers, registries, and authoritative and lineage referencing can be very confusing. An examination of a few simple examples might help to clarify how registration might work for geographic information systems.

18.9.1 Example: Code List Registers

A very simple example is a code list register. One particular information item may have a list of possible attribute values that are allowed. In a land cover classification system a woody growth form type of vegetation land cover might have a leaf type, and there may be a list of allowed broadleaf type shapes. Figure 18.3 shows the BroadLeafShape item as a subtype of the RE_RegisterItem class from ISO 19135-1.

There are 40 values to this code list defined in Annex D of ISO 19144-2 *Geographic information – Land cover meta language* [11]. This annex of ISO 19144-2 is informative and provides a reference content so that nations or organizations may use common definitions in their registers, but

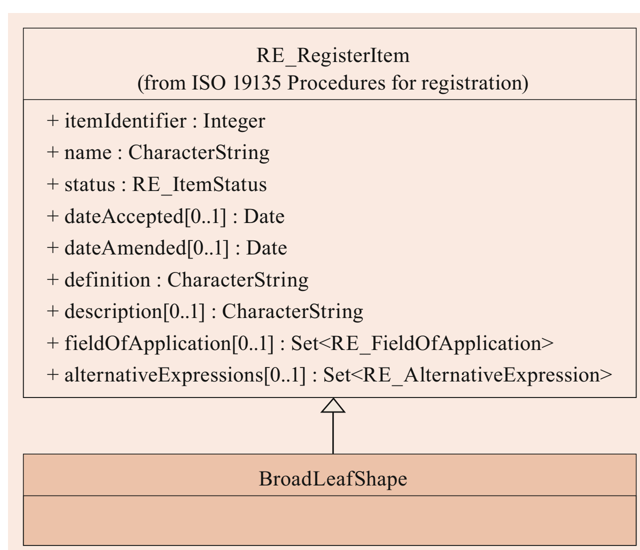


Fig. 18.3 BroadLeadShape item class

Table 18.1 Example implementation of the BroadLeafShape item class as a table

Item identifier	Name	Status	Date accepted	Date amended	Definition	Field of application	Alternative expression
1	Acicular	Valid	1-Mar-10		Slender and pointed, needle-like	Land cover	Acicularis
2	Acuminate	Valid	1-Mar-10		Tapering to a long point	Land cover	Acuminata
3	Aristate	Valid	1-Mar-10		Ending in a stiff, bristle-like point	Land cover	Aristata
4	Bipinnate	Valid	1-Mar-10		Each leaflet also pinnate	Land cover	Bipinnata
5	Cordate	Valid	1-Mar-10		Heart-shaped, stem attaches to cleft	Land cover	Cordata
6	Cuneate	Valid	1-Mar-10		Triangular, stem attaches to point	Land cover	Cuneata
7	Deltoid	Valid	1-Mar-10		Triangular, stem attaches to side	Land cover	Deltoidea

anyone is free to add to the list or change a definition if they need to, using the register management procedures defined in ISO 19135-1. This allows multiple land cover classification systems to exist and be described using the same metalanguage and a federated set of registers. The broadleaf type shapes code list may be extended or modified through the use of a register. Table 18.1 shows a part of the list of 40 values.

18.9.2 Example of Roles for a Feature Register

One of the more important aspects of geographic information to manage through registers is the definition of geographic features. Several of the ISO standards describe how to create feature catalogues (ISO 19110) [12], feature concept dictionaries (ISO 19126), and application schema based on a feature model (ISO 19109) [13]. However, none of the ISO/TC 211 standards actually define a set of features. The system developer must establish a particular feature catalogue for their application. In the very specific area of simulation and modeling, the ISO/IEC JTC1 SC24 *Subcommittee on computer graphics, image processing, and environmental representation*, has defined a specific feature catalogue, ISO 18025 [14]. As stated in the scope of that subcommittee, this is specifically *not* geographic information, but it is related to geographic features in that geographic information is often imported into a simulation system as background information. This feature catalogue is also managed through a register.

The Defence Geospatial Information Working Group (DGIWG), an international standards organization responsible for the interoperability of military geographic information systems, has developed the very widely used DGIWG Feature Data Dictionary (DFDD). This dictionary is managed as a register and contains definitions for a very large number of feature concepts. Based on this dictionary, particular feature catalogues are built for specific data products. This is all done in compliance with the suite of ISO/TC 211 standards. Many organizations, even nonmilitary organizations, use this dictionary as the basis of their national or organization's feature catalogue(s).

This dictionary is too large to show in an example. In this case, the register owner is the DGIWG organization. The submitting organizations are the DGIWG member organizations and liaison organizations (such as the International Hydrographic Organization). There are several registers owned by DGIWG, including the DFDD for features and attributes and specific code lists of allowed values, and metadata including many subregisters, portrayal, and terminology. There are several control bodies in DGIWG. The control body's responsibilities are limited to content control. The register manager does the content maintenance. Sometimes, these two different responsibilities may be vested in a single person or organization. One control body manages the content in the DFDD, and its membership consists of experts nominated by the various DGIWG member nations. There is one registry, consisting of a database with a web interface available [15]. Output to the register user is either in a web browser format or XML data format.

18.10 The EPSG Geodetic Parameter Registry

We now look in greater detail at an example of a register and registry. The EPSG Dataset, or to give it its full title, the EPSG Geodetic Parameter Dataset, is a collection of data that defines several thousand coordinate reference systems used globally, regionally, nationally, and locally. The Dataset is, in ISO 19135-1 terms, the register. It is distributed in several ways, but the prime method for data distribution is through the EPSG registry. The registry is also the mechanism through which the EPSG Dataset is maintained. This section provides an overview of the EPSG registry and its operation.

EPSG stands for European Petroleum Survey Group. In 2005, this group merged into the International Association of Oil and Gas Producers (IOGP), and thus today, the name EPSG is retained only as the name for the Dataset register and registry and associated support documents.

Before we look at the EPSG registry itself, a brief explanation of why it exists is given. Further details can be found in [16].

18.10.1 Coordinates and Coordinate Metadata

Coordinates are used extensively in GIS. They are used in GIS systems for mathematical operations including the description of both absolute and relative positions, the calculation of distances between features and the computation areas. Geographic datasets can only be merged together properly either when the coordinates included in each dataset are consistent between datasets or when the information to convert one set of coordinates to another is available.

However, coordinates are simply numbers. They describe position unambiguously only when what the numbers mean has been defined. It is obviously necessary to know the units and the directions in which the numbers increment, that is, the details of the *coordinate system*. However, to describe position on or near the Earth, it is necessary to define not only the coordinate system, but how that coordinate system is positioned with respect to the Earth itself. A *datum* or *reference frame* can be considered to be the definition of the relationship between a coordinate system and the Earth (or other physical body), and the combination of a coordinate system and a datum is a *coordinate reference system* (CRS) (Chap. 9).

A coordinate reference system definition, therefore, is metadata that makes coordinates unambiguous. However, it is a very particular type of metadata in that it is not optional; without their CRS metadata, coordinates are at best ambiguous and at worst useless.

The EPSG Dataset contains the definitions for coordinate metadata. It is primarily concerned with describing coordinate reference systems used for large and medium-scale mapping and GIS. It also contains definitions of transformations that facilitate the change of coordinates from being referenced to one CRS to being referenced to another. The registry information includes metadata about the CRS definition, for example, the name of the national mapping agency or other authority from which the definition was obtained. This is metadata about metadata, metadata for the CRS record, which itself is coordinate metadata! The working language for the EPSG Dataset and EPSG registry is British English.

18.10.2 Registry Ownership and Management

The EPSG Dataset is owned by IOGP (www.iogp.org). It is managed by the Geodesy Subcommittee of IOGP's Geomatics Committee. This subcommittee also acts as the control body for the dataset content.

Although IOGP owns the registry and through its Geodesy Subcommittee controls the content of the registry, in most cases, IOGP is not the originator of the information. Most coordinate reference systems are defined by national mapping agencies or similar regional, national, or international authorities. The EPSG Dataset simply collates this information from many different sources in a single repository.

The register/registry manager as described in ISO 19135-1 is not necessarily a single person. Management has two principal components: the data content (Register) and the IT system (Registry). For the EPSG register, data maintenance is the responsibility of the IOGP Geodesy Subcommittee, and for the EPSG registry, the IOGP IT Department is responsible for the physical maintenance of the computing infrastructure.

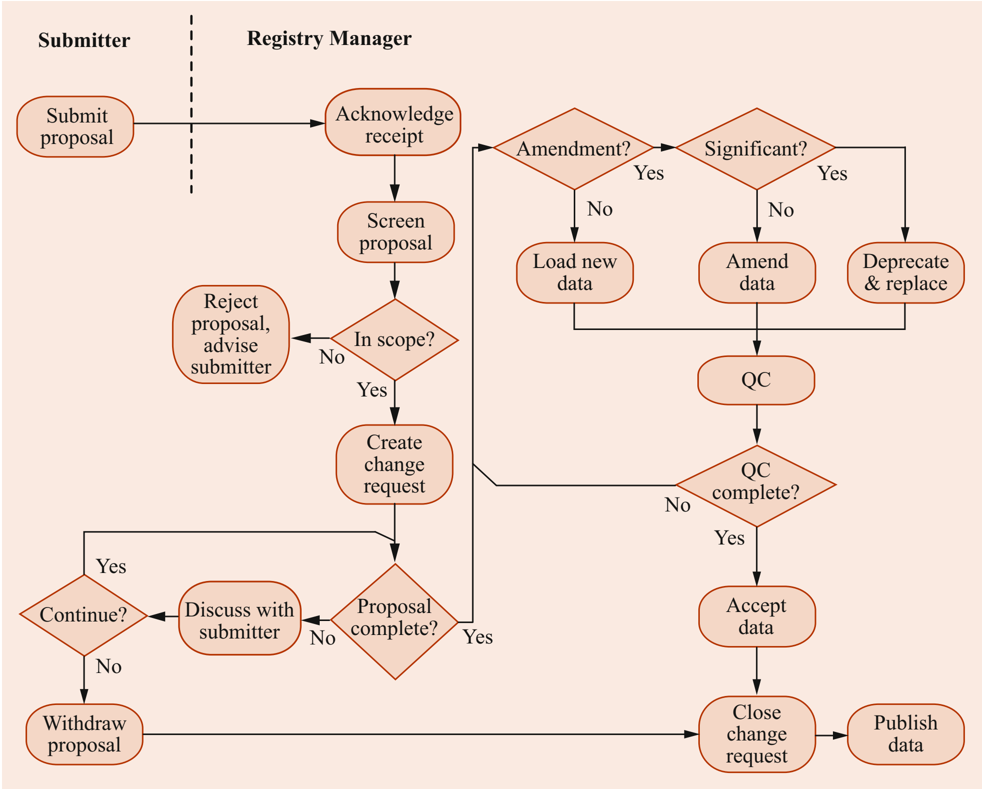
18.10.3 Register Data Management

The flow for processing is shown in Fig. 18.4.

Unlike some registers that have constraints on who can submit proposals for content change, the EPSG register accepts change requests from any interested party. Requests may also be initiated by the register management team. Requests may be for the addition of new data or for the alteration of existing data. Most external requests are received electronically through the EPSG website. Submitters initially receive automated acknowledgement that their submission has reached IOGP. The request is then forwarded to the register manager, where it is quickly screened to check that the proposed data would be within the scope of the register. This initial screening of requests can be time consuming. Some of the communications received are requests for help to understand basic geodetic, cartographic, or GIS concepts and are not anything to do with registry content specifically. These communications will be politely answered but will not be logged as change requests. Other requests may relate to data that is technically within the scope of a register of geodetic parameters but may be so specific to the internal usage of one organization that it does not appear to be of general interest for the data to be promulgated through the EPSG Dataset. In such cases, the submitter will be invited to provide evidence of a wider interest in the data. If he or she is unable to do this, the request will be rejected. When a request clearly falls within the scope of the EPSG Dataset, a change request is created, its identification is assigned, and the submitter is notified of this.

The register data manager, in consultation with other members of the control body, will then review the request in detail. If necessary, he or she will revert to the submitter with questions to clarify aspects of the request. These may be to ensure complete technical understanding, to obtain missing information, or to get agreement on modifications to the proposed data to meet the constraints of the registry data model or its naming conventions. Sometimes, several iterations are required. This process may take days but can take months.

Fig. 18.4 Flow diagram for processing EPSG Register change requests



Code Assignment

Once agreement on content is reached between the register data manager and the submitter, the draft data is loaded to the registry with the status of *pending*. It is at this stage that codes for each entity are assigned. Codes are unique identifiers for items in the registry. For the EPSG register, codes are assigned within a prescribed range through registry software. The data is stored using a hierarchical relational data model, and all components are given codes. Codes are only unique within each entity type; the same code number may be used for different types of entity, for example, CRS and coordinate operation method. The proposal submitter's requests for a particular code assignment cannot be met.

Codes for data in registers may be used as an abbreviated way of giving the full definition to anyone who has access to the register. So by citing, for example, the identifier *urn:ogc:def:crs:EPSG::3456* a data deliverer can rely on the EPSG registry content to give the full coordinate reference system definition. A recipient of the coordinate data and its identifier, who also has access to the EPSG registry, can then put meaning to the code. In this example, the full name of the coordinate reference system is *NAD83(HARN)/Louisiana North (ftUS)*. This is a projected coordinate reference system in use in northern Louisiana in the USA using nonmetric units. A different code is assigned to the description of the same system where meters are the units for the CRS.

Deprecated Data

The EPSG Dataset includes a record status – deprecated – which is an extension beyond ISO 19135-1. The ISO standard provides for data that has been superseded or retired. However, these are lifecycle states that are applicable to processes such as map production, where a second edition of a map will see the withdrawal of the first edition. It is difficult to apply superseded or retired to geodetic parameters. Because coordinate reference systems remain valid for legacy data, they cannot be withdrawn until the legacy data itself is no longer in use, which might be decades after it was published. However, the documentation of geodetic information is done by humans, who are fallible. Unfortunately, errors do enter geodetic registries. The source of the error may be the original information provider or the register manager. Once an error is identified it needs to be corrected. However, to maintain backward compatibility, erroneous data should not be removed from the registry, for even after it has been corrected it may be necessary to use the initial incorrect version to rework calculations that have used the erroneous data. In the EPSG Dataset, if errors are found they are classified as either significant or not significant. For insignificant errors, the record is updated. A significant error is one that would cause a change in the results of a calculation using the data and therefore, usually applies to errors found in parameter values. For these, the record is given a status of *deprecated*, and a replacement record with a new identifier is created. The deprecated record is linked to its replacement.

Public Release of Data
Draft records are then put through the EPSG register manager's quality control (QC) process. This comprises one or more knowledgeable members of the control body critically reviewing the pending new and updated data, as well as considering whether further existing data should additionally be updated. One or more iterations of the QC process may be needed.

After QC has been completed, the data may be released immediately or may be held to be released with other change requests at a later date. The release of data poses a problem for registry managers. The submitter of the request invariably wants an immediate public release of the information, if for no other reason than to be able to cite the code. However, the release process itself utilizes some registry management resources and for the registry manager there are economies of scale in making less frequent but larger releases. Similarly organizations that use the whole registry content to update their own records do not want an irregular yet frequent series of updates but will wish to update on a periodic basis. For example, application developers may synchronize their copy of the registry data only when they have a product release or upgrade. For the EPSG Dataset, historically releases have been made approximately every 6 months. However, for the EPSG online registry (but not other versions of the Dataset), interim releases may be made more frequently than this, at the discretion of the register manager.

18.10.4 Registry Service Interface

The EPSG registry supports a web service interface. The registry uses an application that is conformant to the OASIS (Organization for the Advancement of Structured Information) ebRIM (electronic business Registry Information Model) standard. This distinguishes between data and its metadata. Using ebRIM terminology, the data itself is stored in GML documents as RepositoryItems in a *repository*, and standardized metadata describing the content is stored as RegistryObjects in a *registry*. However, in ebRIM *registry* is used in a very specific way. Both the repository and the registry are internal components of the EPSG Registry.

The GML for an entity may be retrieved through an HTTP GET request.

This will return the GML as given in Table 18.2.

This GML document apparently gives no geodetic information for the projected coordinate reference system. This is because the query returns only the top-level CRS document. It contains xlink references to components such as datum, map projection (conversion), and coordinate system. To build up the full definition of the CRS it is necessary for further calls to be made to each of these components. Some of these may, in turn, require yet more calls. For example, the GML for the projected CRS's base geodetic CRS will include an xlink to the datum, and this in turn will include an xlink to the ellipsoid, after which a further call to obtain the ellipsoid parameter units will be required.

The retrieval of an entity in its entirety starts with first obtaining the ebRIM metadata record for the given CRS. This may be accomplished via a GetRecords request to the *registry*. Secondly, the GML representing the CRS definition must be obtained. The GML document can be acquired by executing a getRepositoryItem request using the identifier of the ebRIM record for the desired CRS, as in the exam-

Table 18.2 GML returned from an HTTP GetRepositoryItem request to the EPSG Registry

```
<ProjectedCRS xmlns="http://www.opengis.net/gml" xmlns:gml="http://www.opengis.net/gml" gml:id="ogp-crs-5179">
<metaDataProperty>
<epsg:CommonMetaData xmlns:epsg="urn:x-ogp:spec:schema-xsd:EPSG:0.1:dataset">
<epsg:type>projected</epsg:type>
<epsg:informationSource>National Geographic Information Institute NGII).</epsg:informationSource>
<epsg:revisionDate>2010-06-28</epsg:revisionDate>
<epsg:changes>
<epsg:changeID xmlns:xlink="http://www.w3.org/1999/xlink"
xlink:href="urn:ogc:def:change-request:EPSG::2010.050"/>
</epsg:changes>
<epsg:show>true</epsg:show>
<epsg:isDeprecated>false</epsg:isDeprecated>
</epsg:CommonMetaData>
</metaDataProperty>
<identifier codeSpace="OGP">urn:ogc:def:crs:EPSG::5179</identifier>
<name>Korea 2000/Unified CS</name>
<remarks>Replaces Korean 1985/Unified CS (CRS code 5178) from 2010.</remarks>
<domainOfValidity xmlns:xlink="http://www.w3.org/1999/xlink" xlink:href="urn:ogc:def:area:EPSG::1135"/>
<scope>Small scale mapping of whole country.</scope>
<conversion xmlns:xlink="http://www.w3.org/1999/xlink" xlink:href="urn:ogc:def:coordinateOperation:EPSG::5100"/>
<baseGeodeticCRS xmlns:xlink="http://www.w3.org/1999/xlink" xlink:href="urn:ogc:def:crs:EPSG::4737"/>
<cartesianCS xmlns:xlink="http://www.w3.org/1999/xlink" xlink:href="urn:ogc:def:cs:EPSG::4530"/>
</ProjectedCRS>
```

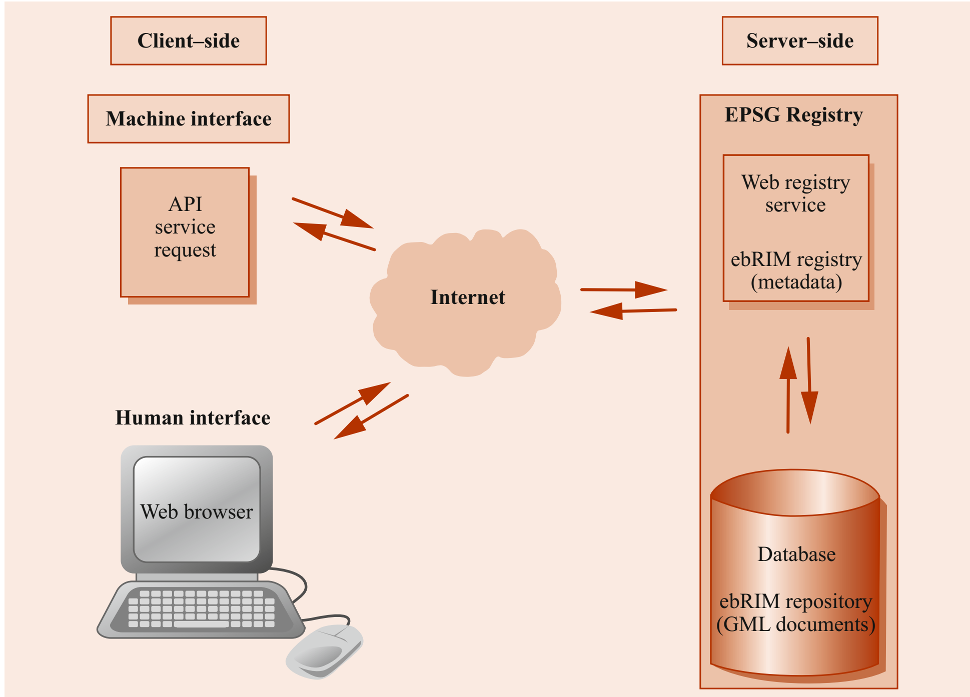


Fig. 18.5 EPSG Registry system components and their interfacing

ple above. Thirdly, once the GML has been obtained, all of the xlink references contained in it must be extracted and resolved. The result after resolving all xlinks for the example CRS is illustrated in the following section. Further details can be found in [17].

It will be clear that this drilling down for component attributes requires a good understanding of the data model used within the registry. The EPSG data model follows that given in ISO 19111 *Geographic information – Referencing by coordinates* [18], but with some extensions for metadata, for example to describe from where the information was sourced. Schema files describing the model are made available. IOGP also provides a guidance note for developers [17]. This details the procedures to be used when constructing API service calls to the EPSG Registry. The EPSG Registry system components and their interfacing is shown in Fig. 18.5.

18.10.5 Registry User Interface

The EPSG Registry includes a graphic user interface (GUI) that allows humans to access the registry content. The user interface is accessible through an Internet browser by using the URL https://epsg.org/search/by-name. It offers a facility to browse content, generate reports, and either export individual entities or the complete EPSG Dataset. For members of the register management team only, it additionally allows for data loading and maintenance. A summary of the facilities available in the EPSG Registry is given in Table 18.3.

Table 18.3 Summary of facilities available to EPSG Registry GUI users

Facility	EPSG Registry user status		
	Guest user	Registered user	Data owner or reviewer
Browse valid data	✓	✓	✓
Report valid data	✓	✓	✓
Export GML for valid data	✓	✓	✓
Browse deprecated data		✓	✓
Report deprecated data		✓	✓
Export GML for deprecated data		✓	✓
Browse pending (unreleased) data			✓
Report pending (unreleased) data			✓
Export GML dictionary of full dataset (valid and deprecated data)		✓	✓
Export SQL script of full dataset (valid and deprecated data)		✓	✓
Export coordinate reference system and coordinate transformation descriptions in accordance with ISO 19162 [19] Well-known Text (WKT)		✓	✓

18

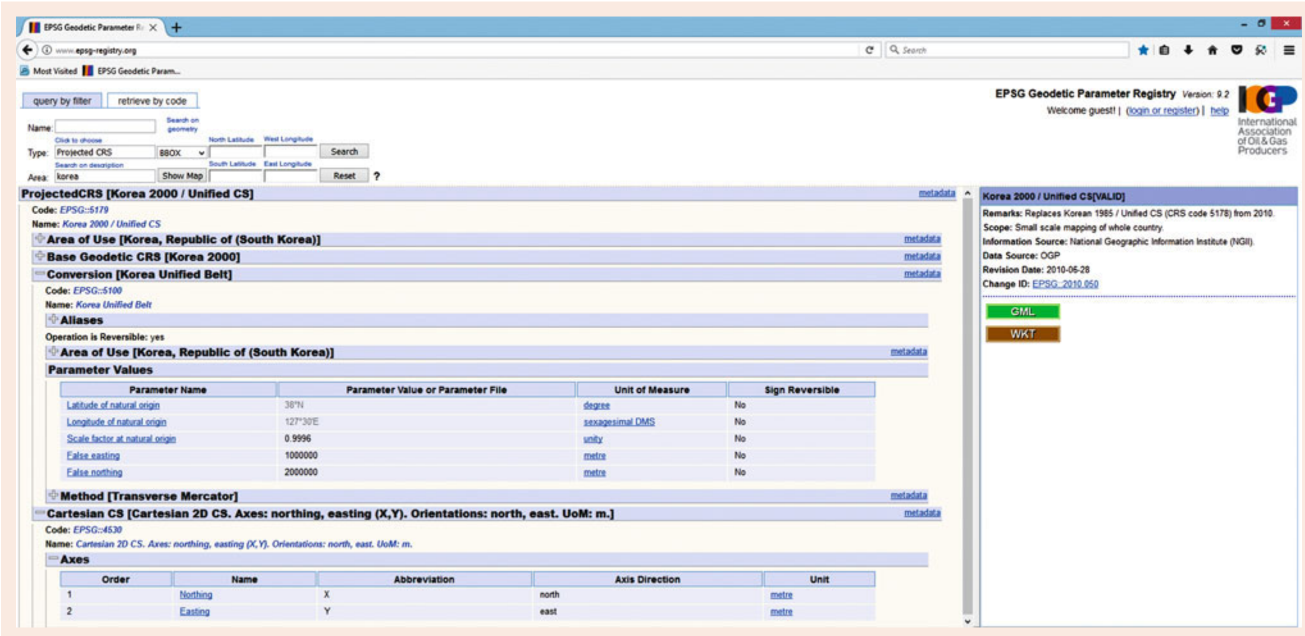
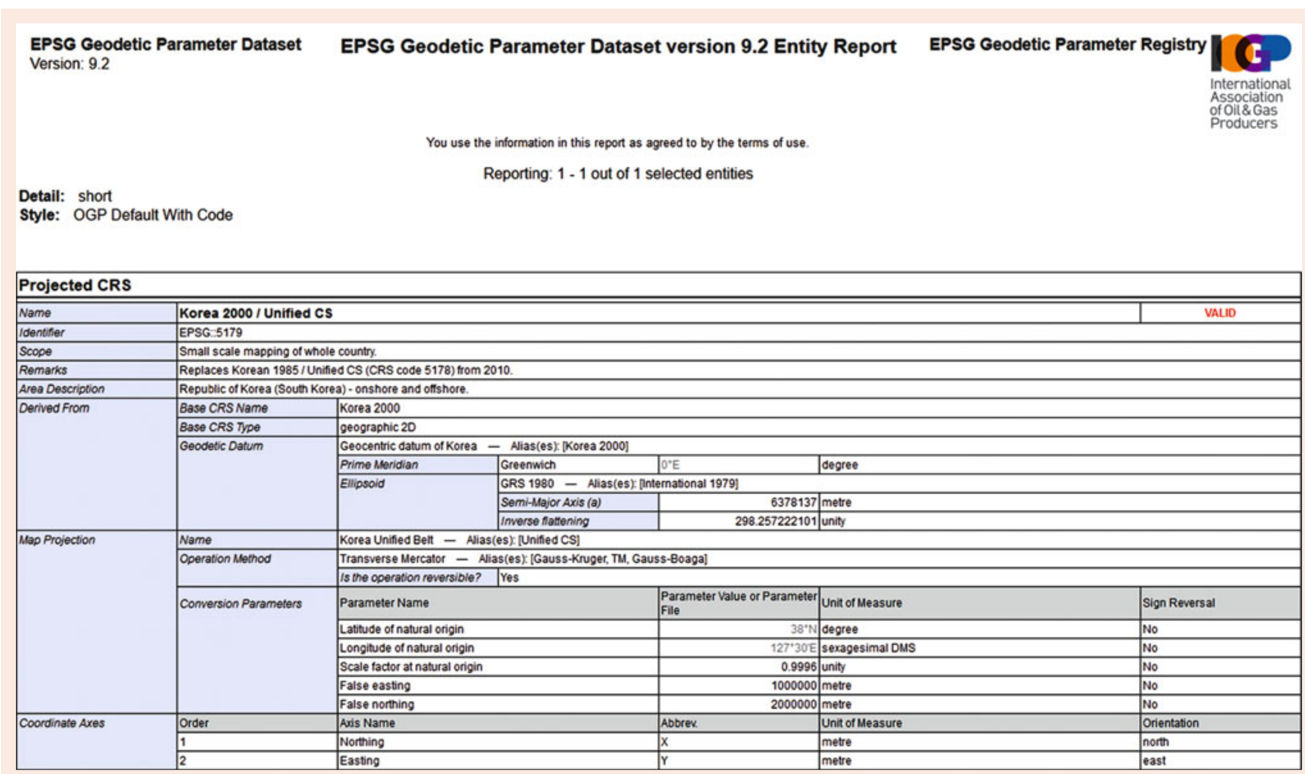


Fig. 18.6 EPSG Registry browse for projected CRSs for Korea



EPSG Geodetic Parameter Dataset
Version: 9.2

EPSG Geodetic Parameter Dataset version 9.2 Entity Report

EPSG Geodetic Parameter Registry

International Association of Oil & Gas Producers

You use the information in this report as agreed to by the terms of use.

Reporting: 1 - 1 out of 1 selected entities

Detail: short
Style: OGP Default With Code

Projected CRS					
Name	**Korea 2000 / Unified CS**				VALID
Identifier	EPSG::5179				
Scope	Small scale mapping of whole country.				
Remarks	Replaces Korean 1985 / Unified CS (CRS code 5178) from 2010.				
Area Description	Republic of Korea (South Korea) - onshore and offshore.				
Derived From	*Base CRS Name*	Korea 2000			
	Base CRS Type	geographic 2D			
	Geodetic Datum	Geocentric datum of Korea — Alias(es): [Korea 2000]			
		Prime Meridian	Greenwich	0°E	degree
		Ellipsoid	GRS 1980 — Alias(es): [International 1979]		
			Semi-Major Axis (a)	6378137	metre
			Inverse flattening	298.257222101	unity
Map Projection	*Name*	Korea Unified Belt — Alias(es): [Unified CS]			
	Operation Method	Transverse Mercator — Alias(es): [Gauss-Kruger, TM, Gauss-Boaga]			
		Is the operation reversible?	Yes		
	Conversion Parameters	Parameter Name	Parameter Value or Parameter File	Unit of Measure	Sign Reversal
		Latitude of natural origin	38°N	degree	No
		Longitude of natural origin	127°30'E	sexagesimal DMS	No
		Scale factor at natural origin	0.9996	unity	No
		False easting	1000000	metre	No
		False northing	2000000	metre	No
Coordinate Axes	Order	Axis Name	Abbrev.	Unit of Measure	Orientation
	1	Northing	X	metre	north
	2	Easting	Y	metre	east

Fig. 18.7 EPSG Registry report for CRS code 5179

The following user statuses are available:

- Guest user
- Registered user
- Data owner or reviewer.

Guest access is the default access level to the registry. Guest users are those users who do not have a registered account in the EPSG Registry or are registered users who are accessing the registry without logging in. Guest users have access only to entities with a status of *Valid*; they do not have access to records with status of *Deprecated* or *Pending*. They can download GML dictionaries of individual records with the status of *Valid* but not of the whole dataset.

Registered users are users who have created a user account through the EPSG Registry user interface and have logged in through the GUI. The EPSG Registry maintains a list of user accounts. Registered users are able to access both *Valid* and (at their option) *Deprecated* records but are not able to access entities with status of *Pending*. Registered users are also able to download GML dictionaries of individual valid or deprecated entities and to use the capability to export the full dataset as a GML dictionary or as an SQL script.

Data owner or data reviewer status is reserved for the register data management team. Only they have access to records with status of *Pending*.

The facilities available through the EPSG Registry user interface to users with differing EPSG Registry status are summarized below.

The user interface allows searching for data within the EPSG Registry by code, name, entity type, geographic extent (area of use textual description), and geographic bounding box (maximum and minimum latitude and longitude). Retrieved results may then be examined at the level of detail required through expansion and contraction of entity details. An example is shown in Fig. 18.6.

Alternatively, reports for retrieved data may be generated. Figure 18.7 shows an example of a report for the same CRS that was used in the service request discussed in Sect. 18.10.4.

References

1. ISO/IEC 11179-3:2003: Information Technology – Metadata Registries (MDR) – Part 3: Registry Metamodel and Basic Attributes. International Standards Organization, Geneva (2003)
2. ISO 19135-1:2015: Geographic Information – Procedures for Item Registration – Part 1: Fundamentals. International Standards Organization, Geneva (2015)
3. ISO 19135:2005: Geographic Information – Procedures for Item Registration. International Standards Organization, Geneva (2005)
4. ISO 19135-2:2012: Geographic Information – Procedures for Item Registration – Part 2: XML Schema Implementation. International Standards Organization, Geneva (2012)
5. ISO 19126:2009: Geographic Information – Feature Concept Dictionaries and Registers. International Standards Organization, Geneva (2009). Revision under development
6. ISO/TS 15000-3:2004: Electronic Business EXtensible Markup Language (ebXML) – Part 3 Registry Information Model (ebRIM) Attributes. International Standards Organization, Geneva (2004)
7. OASIS: OASIS/ebXML Registry Information Model v2.0, April 2002. Organization for the Advancement of Structured Information Standards, Bilerica (2002). http://www.oasis-open.org/committees/regrep/documents/2.0/specs/ebRIM.pdf
8. ISO 19115-1:2014: Geographic Information – Metadata – Part 1: Fundamentals. International Standards Organization, Geneva (2014)
9. ISO 19115-2:2018: Geographic Information – Metadata – Part 2: Extensions for Imagery and Gridded Data. International Standards Organization, Geneva (2018). in press
10. ISO/IEC 19501:2005: Information Technology – Open Distributed Processing – Unified Modeling Language (UML) Version 1.4.2. International Standards Organization, Geneva (2005)
11. ISO 19144-2:2012: Geographic Information – Classification Systems – Land Cover Meta Language (LCML). International Standards Organization, Geneva (2012). in press
12. ISO 19110:20052016: Geographic Information – Methodology for Feature Cataloguing. International Standards Organization, Geneva (2016)
13. ISO 19109:20052015: Geographic Information – Rules for Application Schema. International Standards Organization, Geneva (2015)
14. ISO/IEC 18025:20052014: Information Technology – Environmental Data Coding Specification (EDCS). International Standards Organization, Geneva (2014)
15. DGIWG: *Feature and Attribute Data Registry* (Defence Geospatial Information Working Group, London), https://www.dgiwg.org/FAD/
16. IOGP Guidance Note 7-1, Using the EPSG Geodetic Parameter Dataset, OGP publication number 373-7-1
17. IOGP Guidance Note 7-3, EPSG Geodetic Parameter Registry – Developer Guide, OGP publication number 373-7-3
18. ISO 19111:2018: Geographic Information – Referencing by Coordinates. International Standards Organization, Geneva (2018). in press
19. ISO 19162:2015: Geographic Information – Well-known Text Representation of Coordinate Reference Systems. International Standards Organization, Geneva (2015). Revision under development

Douglas O'Brien C. Douglas O'Brien is an engineering consultant in the field of spatial data engineering and President of IDON Technologies Inc. He was the International Convenor of the ISO/TC 211 WG 6 on Geographic Information – Imagery. He received his Master's degree in Systems Engineering from Carleton University in Ottawa Canada and has been involved in standardization work for over 40 years.

Roger Lott Roger Lott is now retired and was formerly the Chief Surveyor for BP and involved with Geographic Information Systems for over 30 years. Currently Roger Lott is Chairman of the Geodesy Subcommittee of the International Association of Oil and Gas Producers, responsible for maintenance of the EPSG Geodetic Parameter Dataset.

18

Security for Open Distributed Geospatial Information Systems

19

Andreas Matheus

Contents

Abstract

This chapter gives a brief introduction to relevant security requirements and how they can be implemented based on standards for a Web Services and Web-API based approach. It is not the intention to provide individual solutions, as an adequate solution typically depends on many more factors than can be taken under consideration in this chapter. Instead, we like to see this as a starting point from where the reader can follow references to applicable standards for further reading.

Keywords

security · SOA · API · OAuth2 · OpenID Connect · XACML · GeoXACML

A. Matheus (✉)
Secure Dimensions GmbH
Munich, Germany
e-mail: a.m@secure-dimensions.de

19.1 Introduction

Security for geographic information systems (GIS) has gained in importance since Service Oriented Architecture (SOA) enabled the implementation of large, open distributed systems for the creation, processing, viewing, and maintenance of geographic information. Its main characteristic, as specified in [1], is that SOA is a paradigm for organizing and utilizing distributed capabilities that may be under the control of different operators and ownership domains. SOA as an architectural pattern for designing application programming interface(s) (API) that focuses on the separation of responsibility and independence.

As such, SOA APIs cause challenges when implementing effective security functions that take into consideration not only the traditional requirements for installing a GIS in one's own local area network with known and trusted users, but also communication with insecure network segments such as the Internet without knowing which computers and users have access to that network. The traditional paradigm of *we are secure because we have a firewall* no longer holds, as API execution can intrude into an internal system over firewall port 80 via HTTP or port 443 via HTTPS (HTTP over Transport Layer Security/Secure Sockets Layer). However, because there has to be an open port as an essential requirement for participating in a distributed processing system, the question exists of how to properly make one's own system secure and protect it from unauthorized access and prevent attacks that might come in via that open port. Even though it is important to think of all attack vectors when designing an API, it is not the intention of this chapter to elaborate a holistic security approach that encompasses all existing requirements and evaluate all possible options to determine the best solution. Rather, we will address common aspects to provide better understanding of what security means in the context of SOA and APIs and which standards exist to make a geosystem secure for participation in a larger open distributed system.

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_19

When it comes to the decision that *we* intend to participate in a distributed geospatial information system, many questions arise related to security: Are we going to use the traditional Web services or the modern Web API approach? What do we need to do to prevent unauthorized access to the geospatial information and services that we are going to provide? Which potential attacks are we facing, hence which threat models do we need to consider, and can we mitigate or prevent attacks? Can we build a solution based on standards, and which standards are that?

19.1.1 SOA Implementation Options

In service-oriented architectures, any service (component) defines an application programming interface that can be used by other components to communicate/execute the service. Many different implementation approaches for APIs exist. Which one to use highly depends on the overall aspects of the computing environment like scalability, accessibility, availability, as well as robustness and other criteria. For the purpose of this chapter focusing on security, we would like to focus on two common implementation options: (i) Web services and (ii) Web APIs.

Web Services-Based Implementation

When leveraging the traditional Web services approach, the API is a service that can be executed via HTTP. For each service function, a particular input and output message structure encoded in XML must be defined. The entire API can be described using the Web Services Description Language (WSDL) as described in [2]. Therefore, the actual implementation can be based on some HTTP methods like GET and POST and a set of XML messages that define the input and output.

From a security perspective, it is important to note that requirements can be implemented on the actual XML messages independent from the underlying transport protocol. Introducing SOAP [3] and applying WS-Security [4] to the input and output messages, it is possible to implement end-to-end security regarding integrity and confidentiality. Moreover, because the XML processing is part of the application itself, each security feature implemented requires interoperability. This can be achieved by describing the security requirements for the XML (SOAP) messages using the WS-SecurityPolicy standard [5].

Web API-Based Implementation

When leveraging the modern approach to use Web-APIs, one has deliberately tied the implementation to the capabilities of the underlying communication protocol: HTTP. An API endpoint is a URL that allows in combination with a particular HTTP method another component to execute the associated behavior.

From a security perspective, the choice of Web API by itself does not prevent the use of XML messages and, therefore, the option to apply security via WS-Security. However, the modern approach to Web API also implies the use of a more lightweight message structure and the deliberate choice to be more tightly coupled with the capabilities of HTTP [6]. So, instead of sending heavy XML including digital signatures and encrypted SOAP messages, the choice is to use JSON [7] and to attach security with HTTP requests.

The choice of using JSON plus security in HTTP headers is a common approach these days to apply security with Web-APIs and to have the API be accessed by a web application. Even this approach is very tangible for securing one's own system; it introduces additional security challenges and requirements triggered, for example, by the web browser security policy named same origin [8].

19.2 Security Requirements

Before we begin, it is essential to define what we mean by security in the context of this chapter: what it is and is not concerned with. *Security* is described as the characteristic of a system (whether distributed or not) that prevents unwanted, hence unauthorized, actions to be executed on the system itself with potential side effects on information that is accessible via the system. The Trusted Computer System Evaluation Criteria, also known as *Orange Book* states [9]:

> In general, secure systems will control, through use of specific security features, access to information such that only properly authorized individuals, or processes operating on their behalf, will have access to read, write, create, or delete information.

Extending this definition for a single system to a distributed system, which consists of multiple autonomous computers that communicate through a computer network, the communication will not have any influence. This means that the capability of the system to prevent unauthorized access to the information needs to include the communication between the distributed systems.

The typical requirements that exist when securing a distributed system are described in ISO 10181 consisting of

- ISO 10181-1 *Overview* [10]
- ISO 10181-2 *Authentication* [11]
- ISO 10181-3 *Access control* [12]
- ISO 10181-4 *Non-repudiation* [13]
- ISO 10181-5 *Confidentiality* [14]
- ISO 10181-6 *Integrity* [15]
- ISO 10181-7 *Security audit and alarms* [16].

ISO 10181-1 describes the organization of security frameworks, defines relevant security concepts, and describes relationships of the services of the frameworks. To do this it uses security architecture definitions from ISO/IEC 7498-2 [17], such as access control, availability, denial of service, digital signature, and encryption. It also provides other relevant definitions, such as security information, security domain, security policy, trust entities, trust, and trusted third parties, and for the security information, it defines security labels, cryptographic check values, security certificates, and security tokens. In addition, it defines denial of service and availability in such a sense that denial of service cannot always be prevented. In these cases, other security services can be used to detect the lack of availability and allow the application of corrective measures. Annex A of 10181-1 provides an example of protection measures for security certificates and defines the key management framework, as its functions are applicable to any information technology environment where digital signatures and encryption are used.

ISO 10181-2 defines all aspects of authentication in open systems and the relationship with other security functions, such as access control.

ISO 10181-3 defines all aspects of access control in open systems, as it applies to the interactions of user with processes, user with data, process with process, and process with data. It also defines the relationships to other security functionality, such as authentication and audit.

ISO 10181-4 introduces all aspects of nonrepudiation and extends the concepts defined in ISO/IEC 7498-2.

ISO 10181-5 defines confidentiality as a service *to protect information from unauthorized disclosure* in retrieval, transfer, or management.

ISO 10181-6 defines integrity as a property that *data has not been altered or destroyed in an unauthorized manner*. This applies to data in retrieval, transfer or management.

ISO 10181-7 defines the basic concepts of a general model for and identifies relationships between services for security audit and alarms.

When it comes to classified information, and in the geospatial domain, you can find examples for classified information quite easily; additional requirements exist that extend the typical access control requirements where rights are associated to users either directly or by role to ensure the confidentiality of the information and its integrity, including security labels.

Information flow control models such as the Bell–La Padula [18] and Biba models [19] are relevant, as outlined in RFC 1457 [20].

To guarantee the confidentiality of classified information, *The Orange Book* names the Bell–La Padula (information flow control) model [9] that defines secure state, modes of access, and rules that grant/deny access. It ensures that classified information is not flowing from higher classification to lower classification. Therefore, the model is also known for its main purpose: *no read up – no write down*.

The Biba model addresses integrity of information by defining conditions to ensure: *no read down – no write up*.

19.2.1 Thinking About the Threats – What Is the Enemy?

Before thinking of a particular implementation of security aspects, it is essential to think about the relevant, and hence applicable, security requirements. Perhaps it is not always relevant to implement them all. To determine this, the question of which threats potentially exist must be asked. There is a big difference if you consider the *Internet threat model* and/or the *browser threat model* as a relevant cause for any attacks to your system.

With the Internet threat model, it is considered that the communicating end systems can be trusted, but that the communication is unsafe. As defined more precisely in RFC 3552 [21], the attacker has control of the communications channel over which the end systems communicate, and the attacker can read any protocol data on the network and undetectably remove, change, or inject forged information.

In addition to the defined Internet threat model, other threats exist that relate to browsing the Internet that are sometimes listed under the umbrella of the browser threat model. This model considers that the client, the browser application running on an end system, for example, its users are vulnerable to attacks such as phishing, identity theft, etc.

In addition to these general threat models, specific attacks like cross-site request forgery (CSRF or XSRF), also called one-click attack, mainly leverage HTML image tags or JavaScript XMLHttpRequest elements to execute otherwise unauthorized commands as the current user without the user's knowledge. Another form of attack, which is particularly relevant with Web applications implemented in JavaScript, is cross-site scripting (XSS). Here, the aim is to inject malicious JavaScript code as trusted by the application to lever out the Web browser's Same Origin policy.

Without elaborating on this in more detail, it is important to understand which of the listed requirements are important and which standards are applicable to build the solution.

19.2.2 What Is the Web Browser Same Origin Policy?

The Web browser processes content loaded from different Web servers in security sand boxes to prevent malicious website operators to interfere with trusted websites. Different sand boxes exist for different types of content. The Same Origin policy is linked to a sand box for safeguarding the execution of JavaScript initiated network requests. The policy does allow that JavaScript code, contained in a Web page loaded from Web server ‘A’ to load content from the same a Web Server ‘A’ or any other Web server, so long these Web servers are considered “same origin”. The concept of a Web origin is defined in [8], and the detailed protocol for JavaScript initiated network requests to allow cross-origin resource sharing is defined in [22].

19.2.3 Which Requirements Are Geo-Specific?

Requirements stated in ISO 10181-4 are not specific to Geospatial Information Systems (GIS). But geo-specific access conditions must be considered. This has to do with the characteristic of the information: attributes of the information objects as well as the fact that the user can hold geometry information that represents the location, extent, etc. of the object or user. For geospatial data and services or APIs, use cases exist that require the declaration and enforcement of access rights based on the

- Location of the subject
- Geometry of the object (resource)
- Location of the subject and the geometry of the resource
- Topological relations between geometries
- Results of complex processing on geometries.

19.3 Standards for Interoperable Implementation of Security Functions

When it comes to the implementation of security functions, it is a particularly good idea to review existing standards to determine whether there is not (at least) one that can be used. Why? Because many experts have found a keen and practical solution to a problem, and typically software exists, in the form of either libraries or even larger software packages, that has implemented the standard (Chap. 15).

Figure 19.1 provides a first overview of security-related standards that are applicable to secure a distributed geospatial information system based on Web services supporting implementation of the listed requirements. It is worth mentioning that actually one geo-specific specification from the Open Geospatial Consortium (OGC) exists: GeoXACML (Geospatial Extensible Access Control Markup Language). We will elaborate more on GeoXACML in Sect. 19.3.4.

Figure 19.1 is structured such that it categorizes the standards and stacks the layers in a similar way to the Open Systems Interconnection (OSI) model [23].

The Internet Engineering Task Force (IETF) RFCs (request for comments) IPSec (Internet Protocol Security) [24] and TLS/SSL [24] are applicable to actual OSI (Open Systems Interconnection) network layers: IPSec falls into the OSI network layer, and TLS/SSL falls into the transport layer.

The IETF HTTP RFC [6] falls into the OSI application layer, as does SOAP (Simple Object Access Protocol) [3].

As SOAP enables communication using Extensible Markup Language (XML) notation, the next layer above are the XML security standards that contain the W3C recommendations XML digital signature [25] and XML encryption [26].

The next category, message security, is concerned with enabling integrity and confidentiality in XML messages exchanged via SOAP messages. Here the most dominant standard is the OASIS (Organization for the Advancement of Structured Information Standards) WS-Security [4]. As a supplement, one can see the relevance for expressing the requirements that a Web service places on a client to establish communication. WSDL (Web Services Description Language) [2], WS-Policy [27], and WS-SecurityPolicy [5] provide these capabilities.

The next category, concerned with authorization, contains the OASIS XACML [28–30] and the OGC GeoXACML [31–33] standards. An extension to authorization is licensing, which is the next category up. It contains the ISO standard (Mpeg)REL (Rights Expression Language) [34], OMA’s (Outlook Mobile Access) ODRL (Open Digital Rights Language) [35] and content guards XrML (Extensible Rights Markup Language) [36].

Authentication is a cross-layer topic that mainly consists of the IETF RFC for X.509 [37] and OASIS SAML V2 (Security Assertion Markup Language) standard [38–40]. Also, Kerberos [41] and LDAP (Lightweight Directory Access Protocol) [42, 43] for X.500 fall into this category.

Figure 19.2 is structured such that it categorizes the standards and stacks the layers in a similar way to the open systems interconnection (OSI) model [23]. Up to and including the OSI transport layer, the security standards are identical to the ones introduced before. For the OSI application layer, the instance digests in HTTP [44] allow to transmit a checksum for the information send with the HTTP request or response. CORS [22] defines a particular protocol to use HTTP Request and Response headers to overcome the Web browser’s same origin policy.

Even though CORS is a non geo-specific security requirement, it applies to any distributed open system implemented as SOA like a spatial data infrastructure, but in particular using Web APIs and Web applications. One typical example is a Web mapping application that is loaded from Web Server one. Once loaded, the JavaScript executes the mapping application that intends to load maps or perhaps geographic features from other Web servers that are not of the same origin as Web server one. This behavior triggers the Web browser’s same origin policy and it enforces the protocol defined in W3C CORS [22]. Because the Web mapping application will

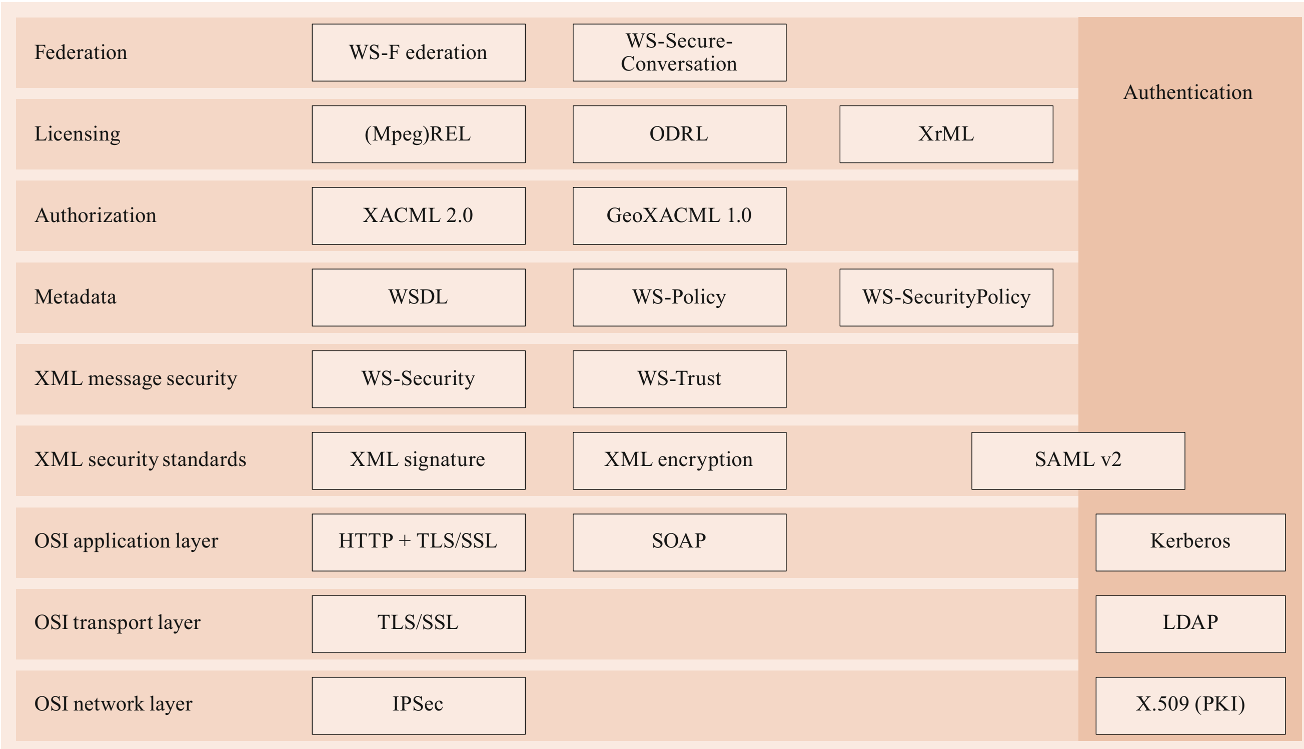


Fig. 19.1 Security standards overview in the context of Web services (subset)

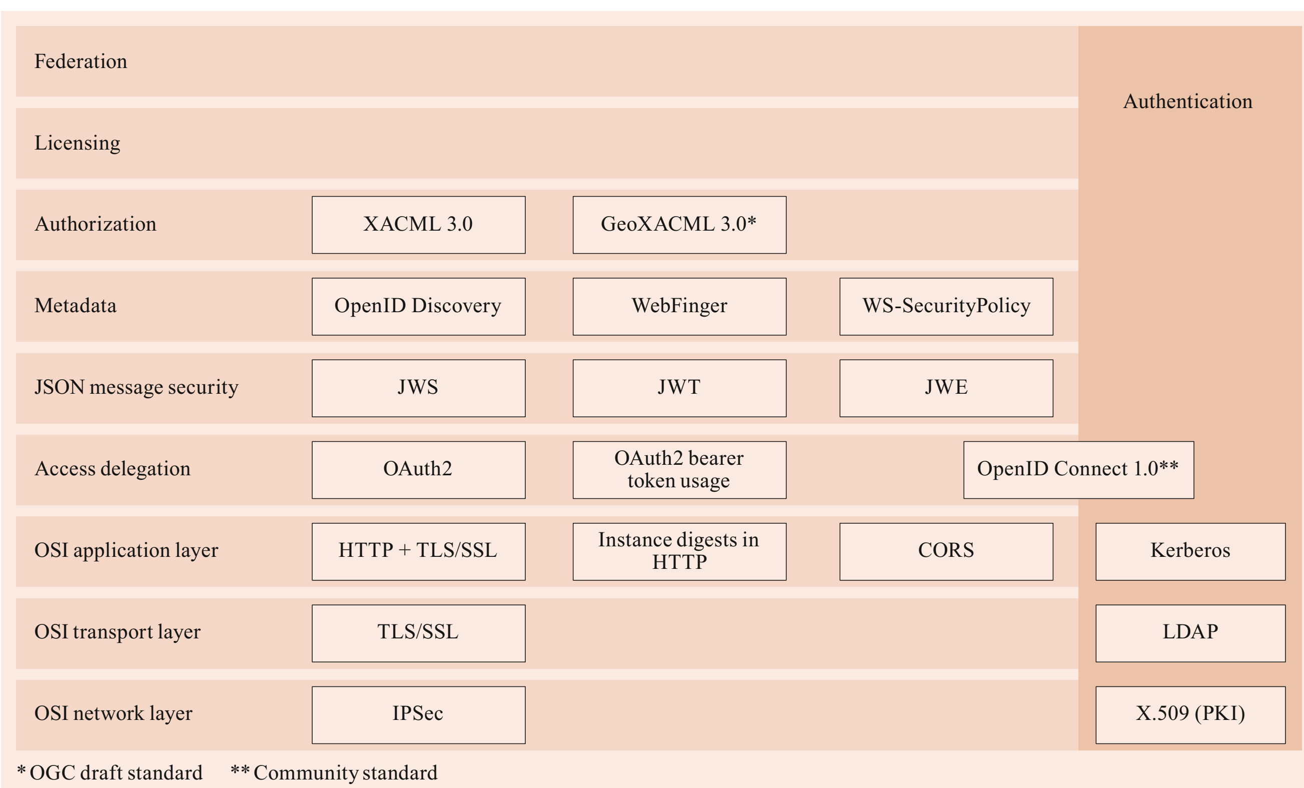


Fig. 19.2 Security standards overview in the context of Web APIs (subset)

only be able to load content from those Web servers that honor the W3C CORS protocol, it is required that each Web API participating in the distributed open system, aka a spatial data infrastructure, must be compliant with W3C CORS.

New from the standards overview as illustrated in Fig. 19.1 is the access delegation layer. The main standards here are OAuth2 [45], the OAuth2 Bearer Token Usage [46], and OAuth2 Token Introspection [47] (not illustrated) RFCs. Also in the layer of Access Delegation is OpenID Connect 1.0 [48] as it is an extension to OAuth2 that bridges to Authentication. It extends the OAuth2 framework with the ability that the Authorization Server can release so called ID tokens that contain user information. It is also possible for Resource Servers to use the specified UserInfo interface to fetch user information associated with an access token.

The JSON Message Security is naturally different from the corresponding layer in Fig. 19.1 as the target is not messages in XML encoding but JSON. The JSON Web Signature (JWS) [49] standard is concerned with the ability to apply a digital signature to arbitrary JSON data. One specialization of JWS is JSON Web Token (JWT) as standardized in [50] that defines how to apply a digital signature to JSON encoded claims. JSON Web Encryption (JWE) as standardized in [51] allows to encrypt the JSON claim. Other standards not illustrated are JSON Web Key (JWK) [52] that is concerned with describing a key in JSON format and JSON Web Algorithm (JWA) [53] that standardizes the registration of cryptographic algorithms with IANA to become usable in JWS and JWE.

The metadata layer contains the standards that are concerned with the description and discovery of endpoints to facilitate the verification of identities of users based on authentication by OAuth2 Authorization Servers as defined in OpenID Discovery [54]. For the purpose of discovery, the OpenID Connect specification leverages the WebFinger [55] standard. The description of Web APIs can be done using OpenAPI [56].

Compared to Fig. 19.1, the authorization layer contains the same standards but with newer versions. In particular, the XACML 3.0 standard [57] enables that authorization request and responses can be encoded in JSON to facilitate better uptake in Web applications. As GeoXACML 3.0 [58], which currently is a draft standard at OGC, is an extension to XACML 3.0, it inherits the JSON encoding for authorization request and responses.

19.3.1 Standards for Implementing Confidentiality and Integrity

Protecting the conversation between two entities can be implemented by leveraging functions from different layers of the OSI reference model; for example, IPSec as a secure extension to Internet Protocol (IP) that resides in layer 4 (the network layer) can be used to encrypt the entire communication between communication end systems. Here, the application itself cannot control how the encryption is done, which is good on the one side, as it takes away the burden from the application programmer to incorporate security functions. A kind of hybrid solution that partially involves the application but still encrypts the entire communication between end systems is TLS/SSL, which can be located in the OSI transport layer. For use cases that require more flexible control over the protection of XML structured communication messages or end-to-end protection, only functions that can be directly controlled by the application and applied to the XML message are feasible.

It is important to note that, for the chaining of Web services, where integrity and, confidentiality span multiple intermediary services, end-to-end protection is required, and therefore, WS-Security-based protection should be applied. Point-to-point protection, as provided by the transport layer, is not sufficient, as information is available in the clear on the intermediary services (Fig. 19.3).

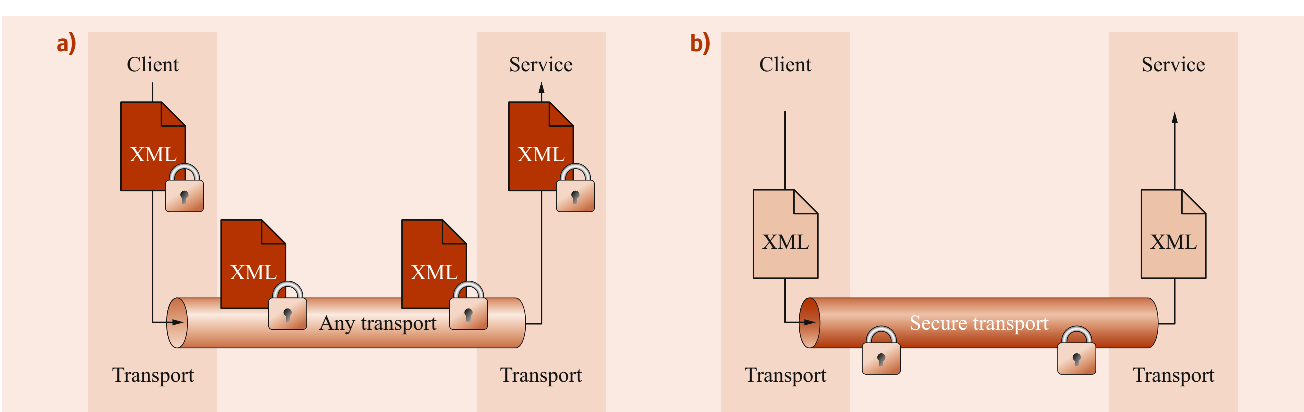


Fig. 19.3 Transport layer (**a**) versus application layer (**b**) integrity/confidentiality

WS-Security is a standard by OASIS that can be associated with the application layer of the OSI reference model. It defines how to use XML digital signature and XML encryption on SOAP messages to ensure confidentiality and/or integrity. Because how and which parts of the message are protected can be controlled by the application in a very flexible manner, WS-Security comes into play, as it defines exact patterns for applying a digital signature to an XML document (or parts of it) and how to encrypt parts of the document and create the relevant metadata for the receiver in XML to undo the encryption or use signed hash values to check integrity. As a full introduction to WS-Security and the related standards would exceed the size of this chapter; the interested reader is encouraged to follow the links given in the References.

With Web-APIs, the information exchange is not based on XML but on the much simpler JSON encoding. Even though HTTPS (HTTP over TLS) is typically used to ensure confidentiality and integrity of transmitted information, it is sometimes inevitable to submit information as part of the URL. In general, but in particular in these cases, the confidentiality and integrity of JSON-encoded information can be established by applying particular IETF specifications, as outlined in Sect. 19.3 and Fig. 19.2.

19.3.2 Standards for Implementing Authentication

The Security Assertion Markup Language V2 is an OASIS standard that first of all specifies a markup language for describing assertions about a subject. SAML2 distinguishes between three different types of assertions:

1. *Authentication assertion*, which provides information about the asserted subject regarding the means by which a subject was authenticated, by whom, and at which time.
2. *Attribute assertion*, which provides information about the characteristics of the asserted subject.
3. *Authorization assertion*, which states that access to a particular resource is permitted or denied for the asserted subject.

SAML2 is one ideal standard to implement authentication in distributed systems, where the user (principal) is known by the identity provider (asserting party) and the protected services are hosted by the service provider (relying party). These two are typically separate entities. To establish secure exchange of assertions concerning the identity of and additional information regarding the user between these parties, SAML2 specifies profiles and bindings. XML digital signatures and XML encryption or both can be applied to guarantee the integrity and or confidentiality of the assertions. The most important profiles are (not ordered)

- *Assertion query and request protocol*, which defines the processing rules for how existing assertions can be queried and the structure of the messages.
- *Authentication request protocol*, which enables the relying party to request assertion statements about the means by which a subject was authenticated.
- *Artifact Resolution Protocol*, which defines how SAML2 artifact references can be exchanged instead of the assertions itself.
- *Name Identifier Management Protocol*, which defines how an asserting party can change the name of an identifier that was previously established and is being used by relying parties.
- *Single Logout Protocol*, which defines a sequence of message exchange with the goal of terminating all existing sessions of the subject with other relying parties in close to real time. However, there is no confirmation message because the logout with all relying parties cannot be guaranteed.
- *Web Browser SSO Profile*, which defines how a Single-Sign-On (SSO) can be established using a (regular) Web browser as the client.
- *Enhanced Client or Proxy (ECP) Profile*, which defines the exchange of request/response messages for a client (not a Web browser) that knows which asserting party to contact.
- *Identity Provider Discovery Profile*, which defines mechanisms by which a relying party can discover which asserting parties a principal uses for the *Web Browser SSO Profile*.

The actual use of one or more of these profiles depends on the deployment environment for the services. To accommodate different characteristics, SAML2 defines multiple bindings for the profiles listed above:

- *SAML2 SOAP binding*, which defines how SAML2 assertions are to be exchanged using SOAP messages and how SOAP header elements are to be used to do so.
- *Reverse SOAP (PAOS) binding*, which describes a mechanism where the client is able to act as a SOAP relay relevant for implementing the ECP profile.
- *HTTP Redirect binding*, which enables the exchange of SAML2 messages as Uniform Resource Locator (URL) parameters. To ensure the length limit of a URL is not exceeded, message encryption is used. This binding is relevant where HTTP user agents of restricted capabilities are involved in the message exchange.
- *HTTP POST binding*, which defines how SAML2 messages can be sent inside an (X)HTML form using base64 encoding.
- *HTTP artifact binding*, which defines how SAML2 request and response messages are exchanged using a ref-

erence – the artifact. This binding is essential for implementing the *artifact resolution profile*.

It is worth mentioning that the applicability of a binding depends on the identified threat model: the Internet threat model allows leveraging of any profile, whereas the browser threat model mandates the artifact profile. The artifact profile relies on a secure *back-channel* between the service and the identity provider to exchange the actual assertion(s). The client just gets hold of the artifact, which is a protected, Internet-wide unique reference to associated assertion(s). However, because the client is missing the keys to set up a trusted back-channel with the identity provider, this profile is safe even if the attacker has prepared the client to intercept and wire-tape the communication. With the browser POST profile, for example, the user assertion(s) is (are) pushed from the identity provider to the service provider through the client. A manipulated client could fetch the assertions and potentially use them to carry out attacks. To prevent this, encrypted and digitally signed SAML2 assertions can be exchanged via the client application.

An alternative approach using a Secure Token Service (STS) is defined in WS-trust [59]. Web Services Trust (WS-trust) is an OASIS standard that defines extensions to WS-Security for managing (issuing, renewing, canceling, validating) security tokens for the purpose of establishing brokered trust relations between Web services of communication partners through the exchange of secured SOAP messages. To support brokered trust, this standard introduces the concept of a STS. To use the STS in an interoperable way, XML message formats are defined. It is important to note that this specification does not define any security token types. It specifies how to deal with them to establish trust between Web services and or clients as not directly trusted communication partners.

19.3.3 Standards for Implementing Access Delegation and User Claims

When a user interacts with services via applications, it might be relevant to allow the application or the service to access protected data owned by the user. In this case, the immediate question becomes, how the user can provide credentials to the application or the service so that access to the user's resources becomes possible without the user disclosing the master credentials, i.e. username and password. The solution is access delegation, which allows the user to delegate a controlled set of rights to the application or service. The standard to achieve this is OAuth2 [45], which is a particular realization of the STS concept [60] adopted for HTTP that standardizes different protocols how access tokens are delivered. The OAuth2 framework defines how the Resource Owner can authorize an Authorization Server to release access tokens to applications so that they can access the user's protected resources hosted at the Resource Server. Any application must be registered with the Authorization Server before it can obtain an access token. This registration process is typically a manual interaction of the application developer and the Authorization Server's registration page. For the simplification of the registration process, additional RFCs [61] and [62] can be leveraged.

As OAuth2 is just concerned with access delegation, applications and services have no information about the acting user; the user acts so to say anonymously. In modern Web applications this is insufficient as personalization and proper salutation is not possible at all.

OpenID Connect [48], which must be considered a community standard because it was not released from a major standardization body, is the extension to the OAuth2 framework that allows the application (or the service) to obtain user information. To better control which pieces of user information can be obtained, OpenID Connect specifies the concept of scopes. A scope like profile, email, or address represent a particular set of user attributes, called claims in OpenID Connect. If compared to SAML2, OpenID Connect defines a simplified version of the Attribute Authority concept. To manage trust with applications and which OpenID Connect scopes an application can request, the application must be registered with the Authorization Server with the scopes it want to use. When executing the application, the user must approve the application to access the user information possible via the authorized scopes. This concept does not exist in SAML2 Attribute Authority.

In the OAuth2 specification, the Resource Server is the passive component that accepts and processes access tokens. The OAuth2 Bearer Token Usage RFC [46] guarantees interoperability and ensures proper processing of access tokens submitted by an application. The OAuth2 Token Introspection [47] defines the interface of an Authorization Server that can be leveraged by the Resource Server to validate access and refresh tokens and to obtain additional metadata for a token.

19.3.4 Standards for Implementing Access Control

The major concern of access control is to prevent unauthorized use or disclosure of protected information. The typical solution is to assign identity rights on objects for particular actions that can be invoked on the object. This is a very challenging task already and becomes even more complicated for a distributed system because harmonization of access rights

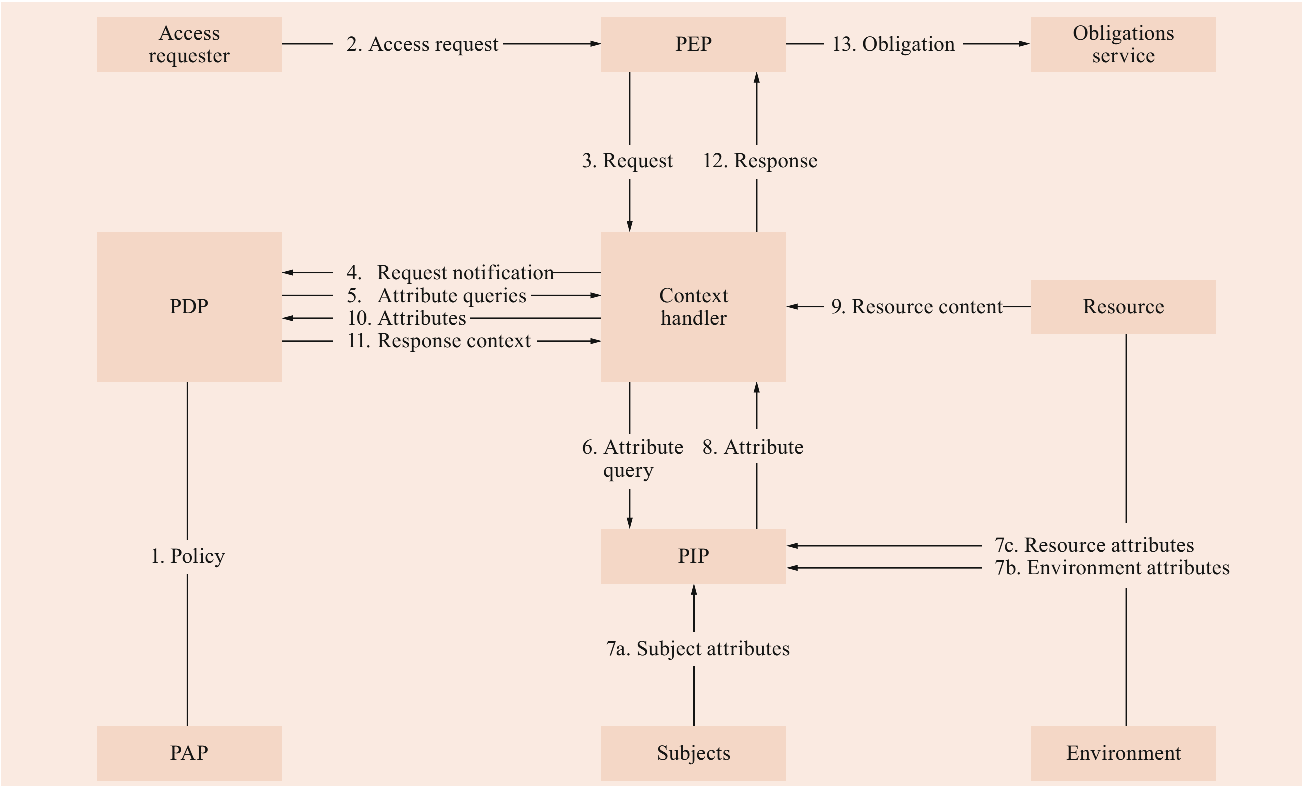


Fig. 19.4 XACML v2 information flow

across jurisdictions requires a language, so that rights declared by one party can be interpreted unambiguously by another involved party.

The Extensible Access Control Markup Language (XACML) by OASIS defines such a language to support the declaration of access rights in XML. It is also possible (of course) to derive authorization decisions based on the rights declared in the policy and an authorization decision request. As the service that derives the decisions (a so-called PDP, Policy Decision Point) can be deployed as an autonomous service, XACML defines the interface and the message format for the XACML authorization decision request and the XACML authorization decision response. XACML V2 [28] mandates the use of XML encoded authorization decision requests and responses, but XACML V3 [57] also supports JSON encoded authorization decision requests and responses [63].

Based on the version of XACML, different profiles exist. The Role-Based Access Control (RBAC) profile defines how to model RBAC0 (pure RBAC) and RBAC1 (role inheritance) [64] in an XACML2 policy. It is important to note that XACML also supports the Bell–La Padula and Biba models to ensure valid information flow control. Through the use of obligations, it is possible to create events for security audit and alarms.

The request to a protected resource is intercepted by the policy enforcement point (PEP). Before the protected resource can be accessed, the PEP involves the context handler to obtain all information relevant to construct a XACML authorization decision request to the Policy Decision Point (PDP). This can involve fetching resource information, and information on the user and the environment through a Policy Information Point (PIP). The PDP, on receiving the authorization decision request, derives an authorization decision based on available policy(ies). The decision is sent back to the PEP, which permits or denies the intercepted request. A decision received from the PDP can optionally contain an obligation, which is to be executed when permitting or denying the request. The Policy Administration Point (PAP) is not involved in runtime processing, as it provides an administrative interface for the creation and maintenance of policies.

As the declaration and enforcement of geo-specific access rights is not supported by XACML, the OGC has drafted a geo-specific extension to XACML 3 [58] and released a standard as the geo-specific extension to XACML 2.0 called Geospatial eXtensible Access Control Markup Language (GeoXACML) 1.0, which builds on top of XACML by using the available extension points. It extends XACML 2.0 by defining the data type *Geometry* and geo-specific functions based on ISO 19125-1 *Geographic information – Sim-*

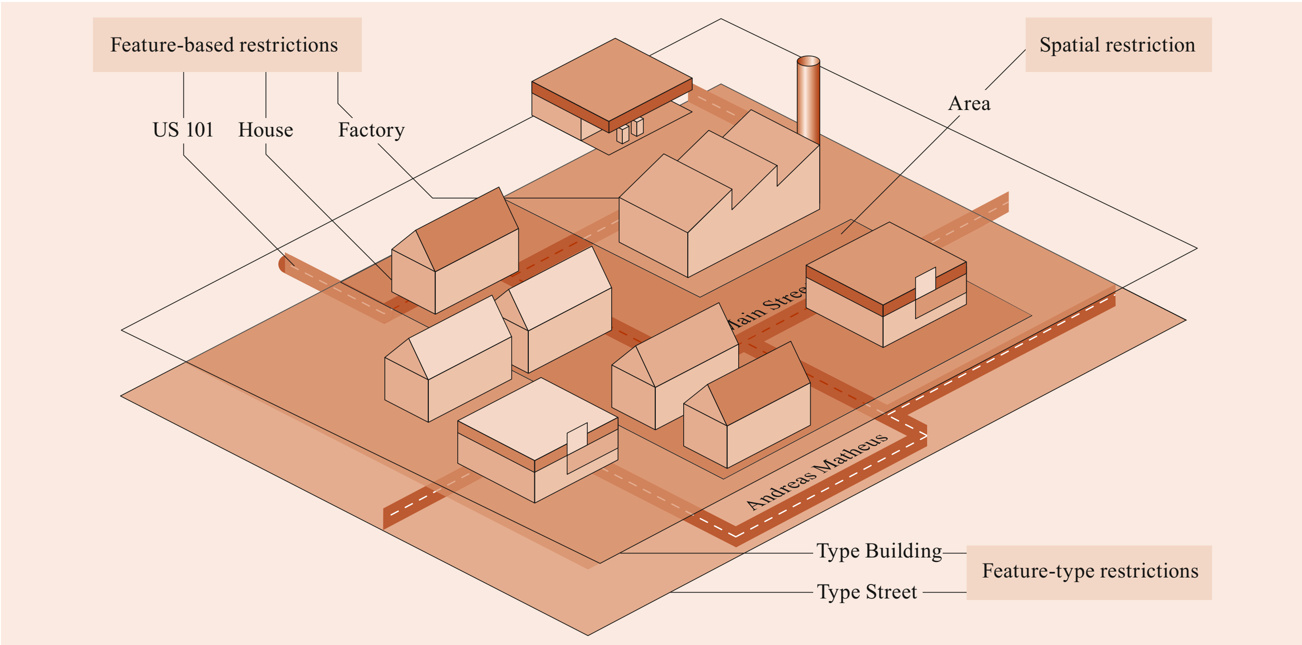


Fig. 19.5 GeoXACML access right example

ple feature access – Part 1: Common architecture, which is identical to OGC document #06-103r3 [65]. The functions allow testing and processing of geometries involved in the process of deriving an authorization decision.

Topological functions allow testing of the topological relation between two geometries; *bag and set functions* allow construction of results or test conditions based on a collection of geometries. Note that the XACML standard defines a bag as an unordered collection of elements with possible duplicates, whereas a set is considered free of duplicates. *Geometric functions* contains constructive and scalar functions for processing new geometries or to request characteristics of a geometry. Finally *conversion functions* (not from ISO 19125-1) support the conversion of length and area values to meters, the mandatory unit of measure.

GeoXACML defines two conformance classes that apply to an implementation of the Policy Decision Point (PDP) as it is a part of the XACML standard informative component diagram (Fig. 19.4). The conformance class BASIC requires a PDP implementation to support the functions listed as topological, bag/set, and conversion functions. The STANDARD implementation of a PDP requires implementation of all functions mandatory for the BASIC conformance class plus the functions listed as geometric functions. In addition, a BASIC or STANDARD implementation must also implement at least one extension (or perhaps all). Currently, the GeoXACML 1.0 core specification is accompanied by two extensions that support the OGC standards GML2 [66] and GML3 [67] encoding of geometries. Because GeoXACML defines an extension to XACML, all of its profiles can be used with GeoXACML too.

Figure 19.5 summarizes the typical capabilities of GeoXACML to control access to a geographic feature.

Rights can be associated with feature types, a particular area, or individual features, as illustrated in Fig. 19.5. As these different types of rights can be combined in any way, one can create very flexible and relevant access policies.

Leveraging the temporal capability from XACML, one can declare and enforce spatial-temporal access restrictions leveraging the Attribute Based Access Control model.

19.4 Summary

Securing a distributed geospatial system mainly involves non geo-specific standards – it requires knowledge of mainstream information technology (IT) to leverage existing standards and implementations in an appropriate way. In this chapter, we introduced an important set of standards covering this subject and associated them with different implementations of SOA: Web Services and Web APIs. The only identified requirement that is geo-specific is access control. Here, an existing standard from the OGC supports the declaration and enforcement of spatio-temporal access rights for geographic information.

References

1. OASIS: Reference Model for Service Oriented Architecture 1.0, OASIS Standard (2006). http://docs.oasis-open.org/soa-rm/v1.0/soa-rm.pdf
2. WSDL: Web Services Description Language (WSDL) 1.1, W3C Note (2001). http://www.w3.org/TR/wsdl
3. SOAP: Simple Object Access Protocol (SOAP), W3C Recommendation, 2nd ed. (2007). http://www.w3.org/TR/soap/
4. Web Services Security: SOAP Message Security 1.1 (WS-Security 2004) – OASIS Standard Specification (2006). http://www.oasis-open.org/committees/download.php/16790/wss-v1.1-spec-os-SOAPMessageSecurity.pdf
5. WS-SecurityPolicy: WS-SecurityPolicy 1.2, OASIS Standard (2007). http://docs.oasis-open.org/ws-sx/WS-Securitypolicy/200702/WS-Securitypolicy-1.2-spec-os.pdf
6. HTTP 1.1: IETF RFC 2616 – Hypertext Transfer Protocol – HTTP/1.1 (1999)
7. Bray, T.: The JavaScript Object Notation (JSON) Data Interchange Format. IETF 8259. (2017)
8. Barth, A.: The Web Origin Concept. IETF 6454. (2011)
9. United States Government Department of the Defense: Trusted Computer System Evaluation Criteria (1985)
10. ISO/IEC 10181-1:1996: Information Technology – Open Systems Interconnection – Security Frameworks for Open Systems: Overview (1996). http://www.iso
11. ISO/IEC 10181-2:1996: Information Technology – Open Systems Interconnection – Security Frameworks for Open Systems: Authentication Framework (1996). http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=18198
12. ISO/IEC 10181-3:1996: Information Technology – Open Systems Interconnection – Security Frameworks for Open Systems: Access Control Framework (1996). http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=18199
13. ISO/IEC 10181-4:1996: Information Technology – Open Systems Interconnection – Security Frameworks for Open Systems: Non-Repudiation Framework (1996). http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=23615
14. ISO/IEC 10181-5:1996: Information Technology – Open Systems Interconnection – Security Frameworks for Open Systems: Confidentiality Framework (1996). http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24329
15. ISO/IEC 10181-6:1996: Information Technology – Open Systems Interconnection – Security Frameworks for Open Systems: Integrity Framework (1996). http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24330
16. ISO/IEC 10181-7:1996: Information Technology – Open Systems Interconnection – Security Frameworks for Open Systems: Security Audit and Alarms Framework (1996). http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=18200
17. ISO 7498-2:1989: Information Processing Systems – Open Systems Interconnection – Basic Reference Model – Part 2: Security Architecture (1989)
18. Bell, D.E., LaPadula, L.J.: Secure Computer Systems: Unified Exposition and Multics Interpretation. MTR-2997 Rev. 1. MITRE, Bedford (1976)
19. Biba, K.J.: Integrity Considerations for Secure Computer Systems. MTR-3153. MITRE, Bedford (1977)
20. IETF RFC 1457: Security Label Framework for the Internet (1993). http://tools.ietf.org/pdf/rfc1457
21. IETF RFC 3552: Guidelines for Writing RFC Text on Security Considerations (2003). http://tools.ietf.org/pdf/rfc3552
22. van Kesteren, A.: Cross-Origin Resource Sharing. W3C Recommendation. (2014)
23. ITU-T: X.200: Information Technology – Open Systems Interconnection – Basic Reference Model: The Basic Model (1994). https://www.itu.int/rec/T-REC-X.200-199407-I
24. IPSec: IP Security – IETF RFC 4301 (2005). http://tools.ietf.org/html/rfc4301. obsoletes RFC 2401 from 1998
25. XML Digital Signature: XML-Signature Syntax and Processing – W3C Recommendation (2002). http://www.w3.org/TR/xmldsig-core/
26. XML Encryption: XML Encryption Syntax and Processing – W3C Recommendation (2002). http://www.w3.org/TR/xmlenc-core/
27. WS-Policy: Web Services Policy 1.5 – Framework, W3C Recommendation (2007). http://www.w3.org/TR/2007/REC-WS-Policy-20070904/
28. XACML 2.0: eXtensible Access Control Markup Language (XACML) Version 2.0, OASIS Standard (2005). http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf
29. XACML RBAC Profile: Core and Hierarchical Role Based Access Control (RBAC) Profile of XACML v2.0, OASIS Standard (2005). http://docs.oasis-open.org/
30. XACML SAML Profile: SAML 2.0 Profile of XACML v2.0, OASIS Standard (2005). http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-saml-profile-spec-os.pdf
31. Open Geospatial Consortium: GeoXACML 1.0: Geospatial eXtensible Access Control Markup Language (GeoXACML) v1.0 (2008). http://portal.opengeospatial.org/files/?artifact_id=25218
32. Open Geospatial Consortium: GeoXACML Extension A: Geospatial eXtensible Access Control Markup Language (GeoXACML) Extension A – GML2 Encoding Version 1.0 (2007). http://portal.opengeospatial.org/files/?artifact_id=25219
33. Open Geospatial Consortium: GeoXACML Extension B: Geospatial eXtensible Access Control Markup Language (GeoXACML) Extension B – GML3 Encoding Version 1.0 (2007). http://portal.opengeospatial.org/files/?artifact_id=25220
34. ISO/IEC 21000-5:2004: Information Technology – Multimedia Framework (MPEG-21) – Part 5: Rights Expression Language (2004). https://www.iso.org/standard/36095.html
35. ODRL: Open Digital Rights Language (ODRL) Version 1.1, W3C Note (2002). http://www.w3.org/TR/odrl/
36. XrML: XrML – eXtensible Rights Markup Language, ContentGuard (2010). http://www.xrml.org/
37. X.509/PKI: IETF, Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile (2008). http://tools.ietf.org/html/rfc5280
38. SAML: Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0, OASIS Standard (2005). http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf
39. SAML-Bindings: Bindings for the OASIS Security Assertion Markup Language (SAML) V2.0, OASIS Standard (2005). http://docs.oasis-open.org/security/saml/v2.0/saml-bindings-2.0-os.pdf
40. SAML-Profiles: Profiles for the OASIS Security Assertion Markup Language (SAML) V2.0, OASIS Standard (2005). http://docs.oasis-open.org/security/saml/v2.0/saml-profiles-2.0-os.pdf
41. Kerberos: Kerberos: The Kerberos Network Authentication Service (V5) (1993). http://tools.ietf.org/html/rfc4120. IETF RFC 4120 (2005) obsoletes 1510
42. LDAP: Lightweight Directory Access Protocol (LDAP): The Protocol – IETF RFC 4511 (2006). http://tools.ietf.org/html/rfc4511
43. IETF: The X.500 String Representation of Standard Attribute Syntaxes: IETF RFC (1993). http://tools.ietf.org/html/rfc1488
44. Mogul, J.: Instance Digests in HTTP. IETF RFC 3230. (2002)
45. Hardt, D.: The OAuth 2.0 Authorization Framework. IETF RFC 6749. (2012)
46. Jones, M.: The OAuth 2.0 Authorization Framework: Bearer Token Usage. IETF RFC 6750. (2012)
47. Richer, J.: OAuth 2.0 Token Introspection. IETF 7662. (2015)

48. OpenID Connect: OpenID Connect Core 1.0 Incorporating Errata Set 1 (2014). http://openid.net/specs/openid-connect-core-1_0.html
49. Jones, M.: JSON Web Signature (JWS). IETF RFC 7515. (2015)
50. Jones, M.: JSON Web Token (JWT). IETF 7519. (2015)
51. Jones, M.: JSON Web Encryption (JWE). IETF 7516. (2015)
52. Jones, M.: JSON Web Key (JWK). IETF 7517. (2015)
53. Jones, M.: JSON Web Algorithms (JWA). IETF 7518. (2015)
54. OpenID Discovery: OpenID Connect Discovery 1.0 Incorporating Errata Set 1 (2014). https://openid.net/specs/openid-connect-discovery-1_0.html
55. Jones, P.: WebFinger. IETF RFC 7033. (2013)
56. OpenAPI: The OpenAPI Specification (2021). https://github.com/OAI/OpenAPI-Specification
57. XACML 3.0: eXtensible Access Control Markup Language (XACML) Version 3.0, OASIS Standard (2013). http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.pdf
58. Open Geospatial Consortium: GeoXACML 3.0 DRAFT Standard: Geospatial eXtensible Access Control Markup Language (GeoXACML) v3.0 (2013). https://portal.opengeospatial.org/files/?artifact_id=55232
59. WS-Trust: WS-Trust 1.3, OASIS Standard (2007). http://docs.oasis-open.org/ws-sx/ws-trust/200512/ws-trust-1.3-os.pdf
60. WS-Trust: WS-Trust 1.4, OASIS Standard Incorporating Approved Errata 01 (2012). http://docs.oasis-open.org/ws-sx/ws-trust/v1.4/errata01/os/ws-trust-1.4-errata01-os-complete.pdf
61. Richer, J.: OAuth 2.0 Dynamic Client Registration Protocol. IETF 7591. (2015)
62. Richer, J.: OAuth 2.0 Dynamic Client Registration Management Protocol. IETF 7592. (2015)
63. XACML 3.0: JSON Profile of XACML 3.0 Version 1.0 (2017). http://docs.oasis-open.org/xacml/xacml-json-http/v1.0/xacml-json-http-v1.0.html
64. Ferraiolo, D.F., Kuhn, D.R.: Role-based access control. In: 15th Natl. Comput. Secur. Conf. pp. 554–563. (1992). http://csrc.nist.gov/groups/SNS/rbac/documents/ferraiolo-kuhn-92.pdf
65. OGC: OGC: OpenGIS Implementation Specification for Geographic Information – Simple Feature Access – Part 1: Common Architecture (2006). http://portal.opengeospatial.org/files/?artifact_id=18241
66. OGC: OpenGIS Geography Markup Language (GML) Implementation Specification, Version 2.1.2 (2002). http://portal.opengeospatial.org/files/?artifact_id=11339
67. OGC: OpenGIS Geography Markup Language (GML) Encoding Standard, Version 3.2.1 (2007). http://portal.opengeospatial.org/files/?artifact_id=20509
68. HTTP 1.1: IETF RFC 7235 – Hypertext Transfer Protocol (HTTP/1.1): Authentication (2014)
69. Lodderstedt, T.: OAuth 2.0 Threat Model and Security Considerations. IETF RFC 6819. (2013)

Andreas Matheus Andreas Matheus received a Dr rer. nat. in Computer Science from the Technische Universität München and is Founder and Managing Director of Secure Dimensions GmbH, a company that develops interoperable security solutions by applying mainstream IT security standards to the geospatial domain. He is an active member of OGC, Chair of the OGC Security Working Group, and a re-elected member of the OGC Architecture Board since 2006. He is Editor of two OGC standards: GeoXACML 1.0 (2008) and OGC Web Services Security 1.0 (2018) as well as editor of the GeoXACML 3.0 draft standard.

Part C
Applications

Maps in the past and geographic information today enjoy a broad range of applications. This handbook selects twelve important applications and provides for each a detailed insight. These range from classical topics such as cadaster, defense, geology, hydrography (marine GIS), and energy & utilities to technologies more recently established in the digital age such as building information modeling (BIM), smart cities, location based services (LBS), GIS in agriculture, transportation, and health. The chapter covering open-source GIS presents nearly 20 individual software products.

The chapter on property cadaster shows two of the many national cadastral systems, first the German system written by Markus Seifert, and second the Dutch system written by Martin Salzmann. Gerhard Joos has written the chapter on defense and explains in it the application of the ISO/TC 211-standards in military applications, namely the requirements of the North Atlantic Treaty Organization (NATO). Kristine Asch together with Jens Klump, Stephen J. Mathers, and Holger Kessler outline applications in geology from historical maps to modern geo-portals. In the chapter marine GIS, Mathias Jonas explains the navigation of ships in a digital environment based on ISO/TC 211-derived comprehensive information systems. William Meehan, Robert Brook, and Jessica Wyland discuss the application of GIS in energy and utilities regarding system design, site planning and distribution to customers.

The chapter on BIM written by Jörg Blankenbach and Ralf Becker guides us into this new topic including its present standards and relationships to neighboring fields. In the chapter location based services (LBS) Haosheng Huang discusses its components and core tasks. Knut Jetlund and Bettina Neuhäuser provide a comprehensive view of GIS in transportation including legal framework, network models, typical databases, data services, and referencing methods. Ralf Bill, Edward Nash, Görres Grenzdörffer, and Jens Wiebensohn discuss the manifold application of GIS in agriculture including precision farming and present developments towards "smart farming". William Davenhall and Christopher Kinabrew start with a brief history of geography and GIS in human health services and illustrate today's relevance of GIS to public health.

The chapter Open-Source Geographic Information Systems was compiled by the chapter editor Peter Löwe. He invited and coordinated fourteen international authors including himself who contributed to this survey of the current range of applications enabled by geospatial open-source software. For this, 18 open-source applications, including core software libraries, desktop/web-GIS, data repositories, data cubes and virtual globes are described. The author's group represents a cross-section of the world's leading developers and stakeholders of open-source geospatial applications and particular FOSS-GIS (Free and Open-Source Software GIS): Álvaro Anguix Alfaro (gvSIG); Andrea Antonello (GeoPaparazzi, JTS, Nasa World Wind); Peter Baumann (rasdaman); Mario Carrera (gvSIG); Kimberly Durante (GeoBlacklight, gdal/ogr); Marco Hugentobler (QGIS); Steve Lime (MapServer, Spatial Libraries: Mapserver Example); Peter Löwe (Introduction, historical development, FOSS licenses, Spatial Libraries: Mapserver Example, PROJ, OSGeo Live); Helena Mitasova (GRASS GIS); Dietmar Müller (GPlates); Markus Neteler (GRASS GIS, PROJ); Jack Reed (GeoServer, Leaflet, GeoBlacklight); Christian Strobl (PostGIS, GeoPython); and Paul Wessel (GMT).

Alan Leidner and George Percivall summarize the impact of digital spatial information technologies on the management of large cities and broadly illustrate the details using New York City as a use case.

Cadastre

20

Markus Seifert and Martin Salzmann

Contents

Abstract

There is an increasing need for standardized cadastral systems to enable involved parties to communicate based on shared vocabulary. This chapter provides an overview of standardization efforts in the field of cadastral systems. There is an increasing need for standardized cadastral systems to avoid reinventing and reimplementing the same functionality over and over again and to enable the parties involved, both within one country and between different countries, to communicate, based on the shared vocabulary. Two approaches are introduced: the Land Administration Domain Model (LADM) and the INSPIRE Data Specification for Cadastral Parcels.

Standardization does not really make sense without looking back at the roots of cadastral systems. Therefore, the German and the Dutch cadastral systems are discussed in this chapter. Both systems vary significantly and can, thus, be seen as examples of the long cadastral history from the early beginning to modern cadastral systems in the digital age where standardization and information modeling are implemented. In The Netherlands, most effort is put into creating technical and semantic interoperability between cadastral and related spatial information sources, starting from the users' needs.

Keywords

cadastre · ISO/TC 211 · INSPIRE · LADM · digital cadastral system · land registry

M. Seifert (✉)
Landesamt für Vermessung und Geoinformation
München, Germany
e-mail: markus.seifert@ldbv.bayern.de

M. Salzmann
Netherland's Cadastre, Land Registry and Mapping Agency
Apeldoorn, The Netherlands
e-mail: martin.salzmann@kadaster.nl

Sufficient cadastral systems have been developed during the last 200 years in most countries in the world. However, due to different approaches, objectives, technology, and societies these cadastral systems have been very heterogeneous until now. One country operates deeds registration, another title registration. Some systems are centralized, others decentralized. Some systems are based on a general boundaries approach, others on fixed boundaries. Some systems have

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_20

a fiscal background, others a legal one. Some are paper-based systems and some are computerized systems. The different implementations (foundations) of the various land administration systems hamper interoperable communication across borders.

However, in all countries, the different cadastral systems are, in principle, largely the same; they are all based on the relationships between people and land linked by (ownership or use) rights and in most countries are influenced by developments in information and communication technology (ICT).

So it was a necessary step to launch harmonization projects in order to decrease this diversity. Two approaches will be explained in the next chapter: the European legislation efforts with the framework directive INSPIRE and the ISO standard ISO 19152 Land Administration Domain Model (LADM) [1]. The main objective is not to standardize the cadastre all over Europe or the whole world but to define a common understanding of cadastral information that can be applied by different users for different pan-European applications.

Those international standards are typically very abstract since they are dealing with the common denominator of the national approaches. For applications on the national level, more detailed information is needed. Considering that there are highly developed cadastral systems in many European countries and around the world, the history, the currents status, and the future steps towards a multipurpose cadastre will be demonstrated with two examples, i.e., the German and the Dutch cadastral systems. While the discussion of the German case puts additional efforts on the new German application schema and the AAA data model as a conceptual basis for the data modeling of any thematic domain model, the Dutch case focuses, among other aspects, on integrating the third dimension and its consideration in the delimitation of rights.

20.1 International Standardization in the Field of Cadastre

The challenges regarding the lack of availability, quality, organization, accessibility, and sharing of spatial information are common to a large number of policies and activities and are experienced across the various levels of public authority in Europe. In order to solve these problems it is necessary to take measures of coordination between the users and providers of spatial information. The Directive 2007/2/EC [2] of the European Parliament and the Council adopted on 14 March 2007 aims at establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) for environmental policies, or policies and activities that have an impact on the environment [2].

20.1.1 The European INSPIRE Data Specification for Cadastral Parcels

INSPIRE is based on the infrastructures for spatial information that are created and maintained by the Member States. To support the establishment of a European infrastructure, *implementing rules* addressing the following components of the infrastructure are being specified: metadata, interoperability of spatial data themes (as described in Annexes I, II, and III of the Directive) and spatial data services, network services and technologies, data and service sharing, and monitoring and reporting procedures. The thematic theme "cadastral parcels" has been considered as reference data and is part of Annex I of the INSPIRE directive. In the context of INSPIRE, cadastral parcels constitute the spatial frame for linking and/or pointing at other information that belongs to a specific thematic field, such as the environment, soil, land use, and many others. Following this approach, the information about the owner, another legal content of the land register, is not relevant in INSPIRE.

For each theme, a data specification has been developed, which contains a UML data model, a GML data transfer format, and further advice for implementation of the data theme, e.g., metadata, data capture rules, and data quality. The INSPIRE data specification on cadastral parcels has been prepared following the participative principle of the consensus building process, which is a general approach of the whole INSPIRE directive development. The data specification was developed by a specific working group, the Thematic Working Group (TWG). The TWG was composed of experts coming from different European Countries and the European Commission, working on a voluntary basis. The specification process took place according to the methodology elaborated for INSPIRE, respecting the requirements and the recommendations of the *INSPIRE Generic Conceptual Model*, which is one of the elements that ensures a coherent approach and cross-theme consistency with other themes in the Directive. The Thematic Working Group for Cadastral Parcels established cooperation with other initiatives within the field, such as the Permanent Committee on Cadastre, EuroGeographics, FIG (International Federation of Surveyors), and ISO/TC 211 responsible for standardization of geographic information.

Scope and Description

The basis of the specification development was the definition of the Directive on cadastral parcels: "areas defined by cadastral parcels or equivalent". In accordance with the particular legal system each member state runs a related register under the responsibility of the government. Such registers are often called a cadastre or, sometimes, a land or other type of registry. Regardless of the name of the system the basic unit of area is the parcel. The generic definition of cadastral parcels has been complemented by the TWG to fit better with user requirements in the following way: the cadastral parcels should be, as much as possible, single areas of Earth surface (land and/or water) under homogenous real property rights and unique ownership, where real property rights and ownership are defined by national laws.

The data model for INSPIRE Cadastral Parcels has been prepared in a way that supports compatibility with the upcoming international standard on LADM. LADM provides a wider context for the INSPIRE cadastral parcels because it includes additional information on rights (bound to national legislation) and owners, which are outside the direct scope of INSPIRE. The next chapter will provide some deeper information about LADM that is currently in revision process at ISO/TC 211.

The wide range of use-cases analyzed by the Thematic Working Group allows meeting the expectations of various user communities in the fields of agriculture, disaster management, soil protection, environmental public right management, public land management, urban planning, utilities, land use, and many others.

The INSPIRE Data Specification for Cadastral Parcels

The core element of the INSPIRE cadastral parcel model is, of course, the cadastral parcel (Fig. 20.1). It is described by some mandatory elements, such as geometry, unique identifier, cadastral reference, and the label of the parcels that supports their identification on printed maps. In case of availability member states are also requested to supply information on the area of the parcel, when the parcel has been created/changed/retired, and the reference point, which is especially useful for visualization.

The management of cadastral parcels in some countries reflects historical subdivision, i.e., intermediate units such as municipalities, sections, districts, parishes, urban and rural blocks, etc. Very often, these units carry information related to all parcels belonging to the same unit, for example information about the accuracy of measurements or the scale of original mapping. In order to be able to refer these units with a common name the notion of cadastral zoning was introduced. EU Member States should decide about the usage of zonings in INSPIRE according to their organizational structures. In addition to carrying metadata, information zonings also support portrayal and data management, especially data search. When the option of using cadastral zonings is selected they have to be supplied with the same properties under the same conditions as stated for cadastral parcels. When several levels of zoning exist in a Member State, it must be ensured that the higher level units are composed of those of the lower level. Cadastral boundaries as separate spatial objects have to be delivered only in the case when information about data accuracy is associated with them.

EU Member States, where national cadastral references are given on basic property units and not on the level of cadastral parcels, have to supply them together with their cadastral reference, unique identifier, area, and the related temporal information.

Interoperability is further supported by the requirement that cadastral parcels have to be published in the ETRS89 or (when applicable) ITRS reference systems, which are commonly used in INSPIRE. When Member States need a common projection system for a cross-border application, the selection must be agreed upon and documented by the interested parties.

For visualization purposes, simple rules for portrayal are given specifying the layout of the borders of cadastral parcels and the cadastral zonings and the cadastral boundaries together with the related labels. These portrayal rules have been defined for several ranges of scales.

The main value of the INSPIRE cadastral parcels model is its simple, yet flexible structure that allows data providers to publish their existing data in the most convenient way. It is also expected that those INSPIRE themes listed in Annex III and related to cadastral parcels (buildings, soil, land use, utility and governmental services, area management/restriction/regulation zones, and reporting units) can reuse and/or further develop the concepts of the current cadastral parcel model.

20.1.2 Land Administration Domain Model (LADM)

LADM is a conceptual data schema. Since land administration is a large field, the focus of LADM is on that part of land administration that is interested in rights, responsibilities and restrictions affecting land (or water), and the corresponding geometrical (spatial) components. LADM provides a reference model that will serve two goals:

- To provide an extensible basis for the development and refinement of efficient and effective land administration systems, based on a model-driven architecture (MDA)
- To enable involved parties, both within one country and between different countries, to communicate, based on the shared vocabulary (that is, an ontology) implied by the model (Chap. 17 Geospatial Semantic Web).

The second goal is relevant for creating standardized information services in a national or international context, where land administration domain semantics have to be shared between regions, or countries, in order to enable the necessary translations. Four considerations during the design of the model were as follows:

- It should cover the common aspects of land administration all over the world.

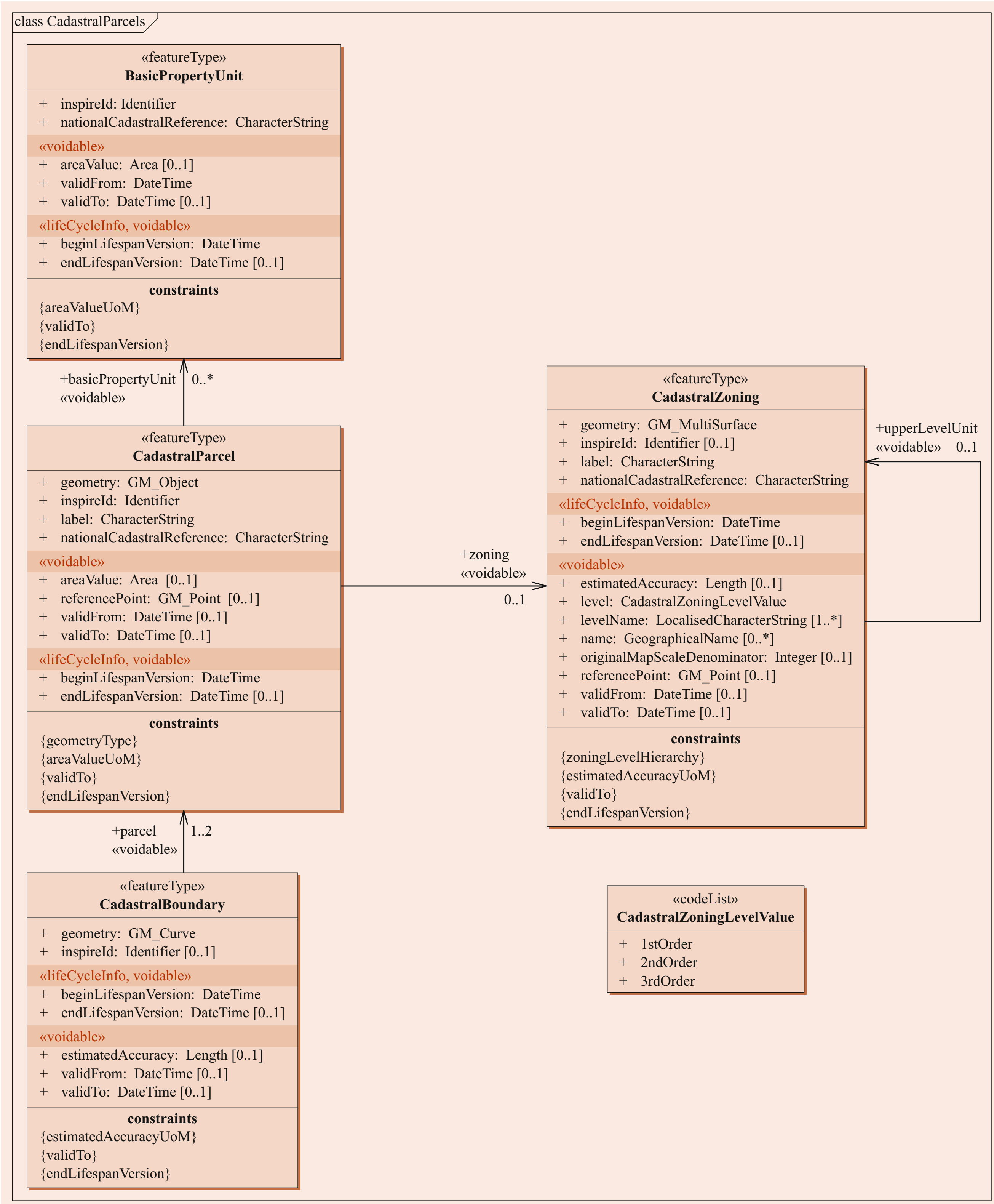


Fig. 20.1 The INSPIRE UML data model for cadastral parcels

- It should be based on the conceptual framework of "Cadastre 2014" of the International Federation of Surveyors (FIG).
- It should be as simple as possible, in order to be useful in practice.
- The spatial aspects follow the ISO/TC 211 conceptual model.

It should be noted that although this is a land administration domain model, it is not intended to be complete for any particular country. It should be expandable, and it is likely that additional attributes, operators, associations, and perhaps new classes, will be needed for a specific region or country.

Organizations are now increasingly confronted with rapid developments in technology (a technology push, Internet, spatial data bases, modeling standards, open systems, GIS), and a growing demand for new services (a market pull, e-governance, sustainable development, electronic conveyance, integration of public data and systems). To meet those requirements, mapping and cadastral administrations use modern data modeling tools. Modeling is a basic tool that facilitates appropriate system development and reengineering and, in addition, forms the basis for meaningful communication between different systems.

Standardization has become a well-known process in the work of land administrations and land registries. In both paper-based systems and computerized systems, standards are required to identify objects, transactions, relationships between objects (e.g., parcels and, more generally, spatial units) and persons (e.g., citizens or subjects, legally speaking, and more generally speaking, parties), classification of land use, land value, map representations of objects, and so on. Computerized systems require further standardization when topology and the identification of single boundaries are introduced. In existing land administrations and land registries, standardization is generally limited to the region, or jurisdiction, where the land administration (including cadastre and/or land registry) is in operation.

Content of LADM

LADM defines a reference covering basic information-related components of land administration (including those over water as well as land and elements above and below the surface of the Earth) and provides an abstract, conceptual schema with five basic packages related to

1. Parties (people and organizations)
2. Basic administrative units, rights, responsibilities, and restrictions (ownership rights)

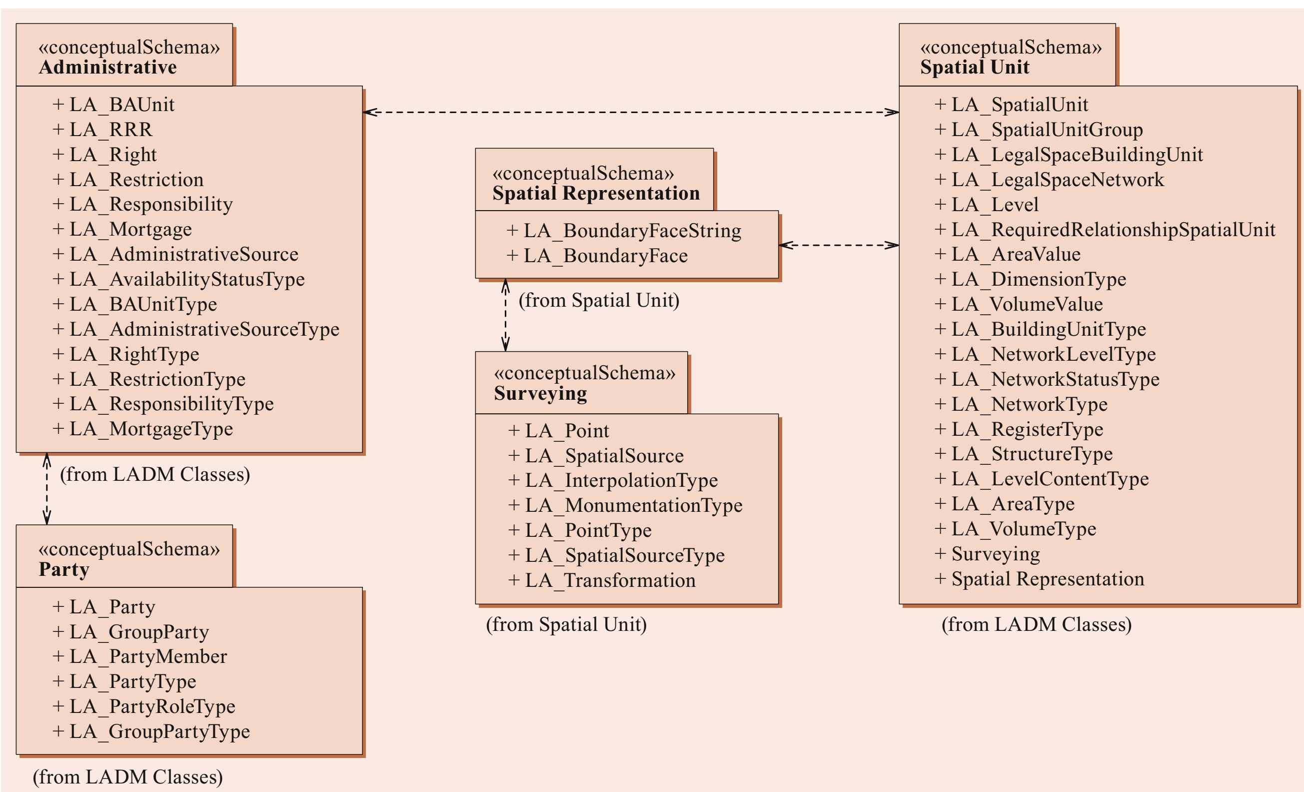


Fig. 20.2 LADM overview of packages (with their respective classes)

3. Spatial units (parcels, buildings, and networks)
4. Spatial sources (surveying)
5. Spatial representations (geometry and topology) (Fig. 20.2).

Additionally, LADM provides a terminology for land administration based on various national and international systems. The terminology allows a shared description of different formal or informal practices and procedures in various jurisdictions. LADM is considered as a basis for national and regional profiles and enables combining land administration information from different sources in a coherent manner.

LADM should be able to accommodate any legal framework. However, legal implications that interfere with (national) land administration laws are outside the scope of LADM. It supports the reform of existing land administration systems at national or other jurisdictional levels through provision of normative and informative annexes that include a number of spatial and legal profiles. These profiles represent specific information arrangements within LADM, offering a variety of modes to organize the geometry and topology of land administration spatial units (e.g., text or point-based, 2-D unstructured, 2-D partition, 3-D full partition) or the legal component (formal real rights, restrictions and responsibilities, including both the private and public originated laws and regulations).

The profiles intermediate between the highly generic and widely applicable nature of LADM and the specification of an individual national land administration model, having its own set of classes, obtained from the LADM generic and abstract classes through a mix of profiling and specialization. The ultimate goal of these profiles is, thus, the support for the implementation of specialized models in existing spatial databases, mainly built around the object-relational or object-oriented paradigms. The provision of such a supporting framework is also fundamental in securing a level of interoperability and semantic translation, once different national specialized models are linked through the common ontology provided by LADM.

Link Between the LADM and INSPIRE Data Specification for Cadastral Parcels

The INSPIRE Directive requires taking existing standards into account (article 7 of the Directive). Once adopted, the ISO 19152 standard should be taken into account if there are requirements and consensus to extend the Data Specification for Cadastral Parcels. In the case of LADM, there was an excellent opportunity, as both INSPIRE Cadastral Parcel Data Specification and ISO standard LADM were under development at the same time. Through joint work between the INSPIRE TWG Cadastral Parcel and the LADM Project Team this was achieved. This ensured consistency between INSPIRE and LADM and resulted in a matching of concepts and compatible definitions of common concepts. Of course, it must be remembered that there are differences in scope and targeted application areas; e.g., INSPIRE has a strong focus on environmental users who mainly need cadastral parcels as a reference for other thematic data, while LADM has a multipurpose character (also supporting legal security, taxation, valuation, planning, etc.). LADM supports both data producers and data users in these various application areas. Also, LADM has harmonization solutions for rights and owners of 3-D cadastral objects (such as building or network reserves), which are currently also outside the scope of INSPIRE Cadastral Parcels. However, through intensive cooperation, it has now been made possible for a European country to be compliant both with INSPIRE and with LADM. Further, it will be made possible through the use of LADM to extend INSPIRE specifications in future if there are requirements and consensus to do so.

In order to "proof" the compatibility, the ISO 19152 LADM document [1, Annex G] includes a LADM-based version of INSPIRE Cadastral Parcels, showing that the INSPIRE development fits within LADM and that there are no inconsistencies [1]. The INSPIRE cadastral parcels model can be derived from LADM. In the INSPIRE context, four classes are relevant:

1. LA_Parcel as the basis for CadastralParcel
2. LA_LAUnit as the basis for BasicPropertyUnit
3. LA_FaceString as the basis for CadastralBoundary
4. LA_SpatialUnitSet as the basis for CadastralZoning.

The LADM attributes inherited by INSPIRE can have a more specific data type or cardinality in INSPIRE (compared to LADM). This has been included in the ISO standard 19152, Annex G. This implies that an optional LADM attribute [0..1], might not occur at all in INSPIRE, as the cardinality can be set to 0; e.g., nationalVolume. This also implies that an optional LADM attribute [0..1], might be an obligatory attribute in INSPIRE; e.g., label. Further, INSPIRE-specific attributes are added to the different classes [2].

20.2 Cadastre and Land Register in Germany

20.2.1 History of the German Cadastre

At the beginning of the nineteenth century, the cadastral systems were established mainly for taxation purposes. In some of the western provinces of Germany, the establishment of cadastral systems for taxation purposes are based on the cadastres established by Napoleon in the occupied areas left of the river Rhine. Although the main purpose for establishing a cadastre was taxation of land, the idea of using maps and

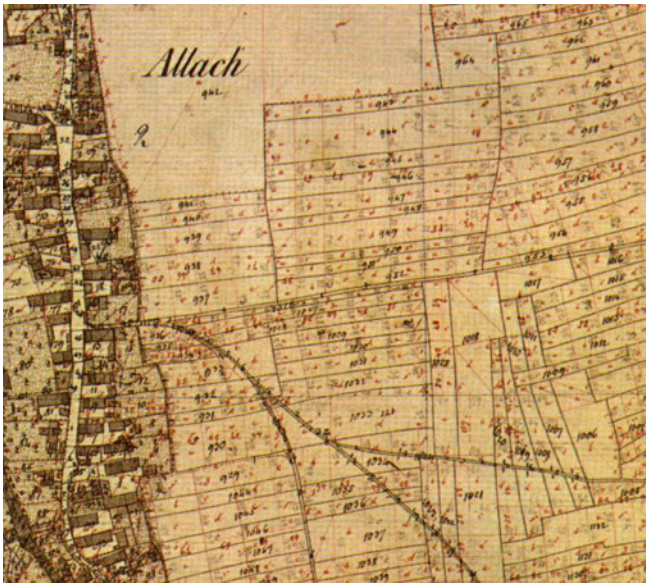


Fig. 20.3 An early cadastral sketch

records for further purposes of governmental activities was implemented in the cadastral systems from the beginning.

The creation and maintenance of the cadastre in Germany was and is the task of the federal states ("Länder").

With the example of Bavaria, the following describes the beginning of the cadastral register in Germany. The origin of the cadastre goes back to the beginning of the nineteenth century, when there was an urgent military need for topographical maps covering the whole territory of the country. So, the "Topographical Office" was founded by the French during the Napoleonic wars in 1801. At the same time, taxation of the land was one of the most important sources of income for the state. However, there were some 114 different and incompatible systems for ground taxation. The state administration realized that the only way to get rid of this diversity was a systematically surveying of all parcels based on a homogenous coordinate reference system. Again in 1801, the Bureau de Cadastre was established as part of the Topographical Bureau. That was the starting point of the creation of cadastral maps in the scale 1 : 5000, 1 : 2500, and 1 : 1000 for the whole country. Figure 20.3 shows an example of a 1 : 5000 cadastral map. For this surveying, a new coordinate reference system (Soldner coordinates) was introduced based on a new triangulation network. Two baselines were measured with contemporary techniques (Figs. 20.4–20.6).

The start and the end point of baseline were marked with stable stone pyramids, which are still in place today. The accuracy of the measurement was amazing. A GPS-based survey of the 28 km long baseline showed an accuracy of about 20 cm.

After the German Reich was founded in 1871, the need for standardization of private law was evident. Since 1 January, 1900 common private law exists for the entire country. In this context, the land registration system for the whole country was established. This land registration system (in German: "Grundbuch") collects all rights of ownership and other rights on land and buildings. With the establishment of this system the importance of a sufficiently working cadastral system grew very fast. The description of land parcels (parcel identifier and cadastral maps) became the official and legal register of parcels as a part of the land register. Cadastre developed from a system for taxation of land to a register to guarantee the right of land tenure.

To sum up, the main objectives of the cadastre in Germany are:

- A legal base for a standardized and fair taxation of land
- A register to back up the property of land as a first step towards a land register
- Support of different official and economical tasks like the real estate market, planning, and the creation of topographical maps. Until now, we have used the term multipurpose cadastre, supporting many different applications. Figure 20.6 shows a very early example of a multipurpose cadastre containing buildings, land use information, and addresses. After 1934, the results of official soil assessment were recorded in the cadastre. This was the first step in the direction of a multipurpose cadastre.

Fig. 20.4 The baseline between the start and end pyramids surveyed with contemporary methods

Fig. 20.5 An early cadastral sketch (nineteenth century)

Fig. 20.6 Early example of a multipurpose cadastre containing buildings, land use information, and addresses

The Federal Republic of Germany is situated in the heart of Europe with nine neighbor countries. Germany is a link between east and west and between Scandinavia and the Mediterranean. It covers an area of about 357 000 km^2. For centuries, the map of Germany showed a patchwork of smaller and bigger independent kingdoms and principalities. The most important changes in the modern age resulted from the Napoleonic wars at the beginning of the nineteenth century, the Austro-Prussian War of 1866, the First and the Second World Wars.

The political and administrative principles in the Federal Republic of Germany are fixed in the constitution, the basic law since 1949. The Federal Republic consists of 16 Länder (states). These states are not just provinces, but states endowed with their own powers. Each has a constitution, which must be consistent with the republican, democratic, and social principles embodied in the basic law. Subject to these conditions, the states can shape their constitutions as they see fit. The basic law determines the powers of the states and in which areas states are allowed to make their own laws. Legislation and administration in the field of cadastral law is in the hands of the states.

Bavaria is one of the 16 states ("Länder") in Germany. It is located in the southern part of Germany. Bavaria contains an area of some 70 000 km^2 and has a population of about 11 million.

The cadastral register contains approximately 10 million parcels and 3.5 million buildings. Each year 55 000 parcels and 160 000 buildings are surveyed.

Traditionally, in Germany the cadastre is separated from the real estate register that is the responsibility of the Ministry of Justice. The establishment and maintenance of the cadastre in Germany lies in the responsibility of the states (Länder). In the states, different ministries are responsible for the cadastre; e.g., in Bavaria, the Ministry of Finance. The cadastre describes all parcels of the area of the national territory. All information concerning boundaries, geometric extension, land use, and location of parcels is determined in cadastral maps and cadastral registers. On the other hand, the land register keeps information about ownership, leasehold, and other legal appointments. The land register and the cadastre depend on each other.

The security of land tenure in Germany has a tradition of some centuries. The forms of land registration in former centuries had great variety because of the great number of independent states in the territory of the present Federal Republic of Germany. The basic laws were quite different as well, because after the end of the 30 Year's War in 1648 states like Prussia or Saxony were completely independent.

20.2.2 Related Organizational Details

With respect to the Federal Constitution of Germany the responsibility for legislation in the field of cadastres is in the hands of the states. The states have passed various, basically uniform, laws in the field of surveying. However, the ministry responsible for surveying varies from state to state. In the majority of German states, surveying and cadastral services have a three-tier structure:

- **Supreme surveying and cadastral authority** (state government ministry): dealing with principles, preparing legislation, and issues administrative rules and guidelines for a uniform implementation of the tasks as the top supervisory body.

- **Higher surveying and cadastral authority:** duties at this level are carried out by the governors' departments (administrative subdivisions of the state governments) covering the territory of several shires and unincorporated cities. They supervise local cadastral and surveying agencies, as well as licensed surveyors, issue general and specific directives, and provide assistance wherever the local services need it.
- **Lower surveying and cadastral agencies**: their jurisdiction covers the territory of a shire or an unincorporated city. They are tasked to maintain and update the real property cadastre. The organization of a cadastral agency varies when these offices have taken over more communal tasks like land valuation, planning, etc.

In all states except Bavaria, licensed surveyors are mandated to perform the cadastral surveys. Notaries are involved in the legal part of the business, authenticating all kinds of contracts associated with buying and selling of land, mortgages, etc.

Each state has its own licensing law for private, licensed surveyors. The preconditions for getting a license for one specific state are a university diploma for surveying and an education in becoming a civil servant with a final examination.

20.2.3 The Land Register in Germany

The land register contains real estate properties. A real estate property is a legally defined spatial part of the Earth's surface, which is posted alone on a special sheet of the land register or under a special title together with others at a common sheet/list of titles of the land register. A real estate property may consist of several parcels.

Thus, a real estate property

- Always has a separate title in the list of titles in the land register
- May consist of several parcels
- May represent a noncontiguous area.

The cadastral register contains parcels. A parcel (in the case of cadastre) is a restricted spatial, not separated surface of the Earth. A parcel is a unit of description and cartographic representation of surfaces in the cadastre.

The constitution of the Federal Republic of Germany accords responsibility for legislation around the land register to the Federal Republic, whereas the states make the laws concerning the property cadastre. The land registration offices are part of the administration of justice in the 16 German states. The jurisdiction of the land registry officials, the form of establishing and maintaining the land register, are manifested in the legal regulation for the land register (GBO – Grundbuchordnung) and in additional orders of the Federal Minister of Justice. The regulations of the organization of the land registry are given by the states' ministries of justice. The land registries are part of the local courts. They are responsible for the land registration of the properties of land in their district. These districts are manifested by law. Land registry officials are

- The **district court judge** as a land register judge
- The **administrators of justice**
- The **nominated official** of the land registration office
- The **official for certifications** of the land registration office.

The administrator of justice is responsible for all external business of the land register. He is able to decide independently, his decisions are regulated and restricted only by law. The official for certifications has to register the contents of the administrator of justice in the register and countersign the records in the register.

The **land register** is characterized by two important legal principles:

- Changes of rights to land are not legally valid before being registered in the land register.
- Until otherwise proven, the correctness of all titles recorded on the register is assumed. The land register enjoys "public faith"; in other words, the details registered may be presumed correct by anyone acquiring a legal title with regard to property unless an appeal against the correctness is recorded or the person acquiring the title is aware of the entry being incorrect.

Registrations therefore play an important legal role. They are in force until evidence of the reverse. So are some contents of the cadastre, like the parcel-identifier in maps and records. The contents of the land register are based on private contracts certified by public notaries. It is not a deed system, but the contracts are stored in the files belonging to each record. Registrations are done only by application. So normally the land registry does not become active by itself. Rights with respect to land being able to be registered in the land register are:

- Ownership
- Shared ownership
- Ownership of a building without ownership of the land (long-term lease)
- Full ownership of an apartment in combination with part-ownership of the land

- Easements as a limited right of an owner of one parcel to use or prevent use of some kind of a neighboring parcel. The right is connected to a parcel, like rights-of-way
- Easement as a limited right of a person or the community, like the right of the municipality to build a gas pipe on private ground
- Mortgages in different forms to secure loans
- Right of first refusal on a parcel of land

There are two ways to register land tenure:

- The real folio, in which each parcel is registered in a special folio (sheet of the land register)
- The personal folio, in which all parcels of an owner are registered in one folio.

The personal folio is the common form for the time being. Generally, each parcel has to be registered in the land register, except parcels in state ownership, which are not part of a normal land market, e.g., streets. Each sheet of the land register has a structure that is described as follows:

1. **Title**
   - Responsible district court
   - District
   - Number of folio.
2. **List of properties**
   - Numbers of properties
   - Numbers of parcels (a property may be separated into a number of parcels)
   - Description of parcels (area, land use, location)
   - Rights of the owner to other parcels (leaseholds)
   - Type of shared ownership.
3. **Part I**
   - Name of the landowners
   - Number of properties.
4. **Part II**
   - All forms of easement, protests, and other restrictions
   - Ranking of rights.
5. **Part III**
   - All forms of mortgages.

In addition to this official register there are files containing contracts and other documents related to the folio. Currently, an object-oriented land registry system has almost been completed.

Not the cadastre but the land register shows the **legal status** of all real properties. In its documentation and publication role, it works as the statutory basis for property security, in particular to ensure the unique status of ownership and other titles, as well as for mortgage loans.

In Germany, the cadastre is defined as the official register of all parcels and buildings in a state, in which all parcels are described with graphical and textual data. Parcels are areas defined by cadastral registers. The cadastral register is designed to show the **de facto status** of real estate properties. As far as legal property titles are concerned, the cadastre shows their scope and the part of the surface to which they extend. All relevant facts, such as designation, location, size, and use, plus the boundaries as surveyed by authorized government agencies and licensed surveyors are described. The cadastre is the only register in which **all** parcels and buildings in a state are described. In addition, it contains further information, like the results of the official soil assessment. With some parts of its contents, the cadastre enjoys the "public faith" of the land register, like parcel identifiers in maps and records. That means that people can rely on the correctness of the entries even if they are wrong.

Cadastre became the basic element of all kinds of geographical information systems and is part of spatial data infrastructure (SDI) projects on local, state, and national levels. In Germany, there exists only one type of cadastre covering the complete territory with all kinds of landowners (private and state) and all kinds of land use (urban, rural, forests etc.).

20.3 The German Cadastral System in the Digital Age

The property **cadastre** in Germany is a parcel-based system, i.e., information is geographically referenced to unique, well-defined units of land. These units are defined by formal boundaries marking the extent of the land. Each parcel is given a unique parcel number. The cadastre, based on cadastral surveying, shows the division of the ground in form of parcels and contains information about the de facto status of the property (location, size, use, etc.). The land register contains the describing parts of the cadastre. Therefore, only both systems in combination are able to give a complete overview about legal and de facto land tenure. Both registers are constantly updated and kept consistent.

Real estate cadastre is the official register of land plots for the proof of ownership in the land register. However, this does not mean that the cadastre builds a register in the sense of [3]. The position of the freehold property is presented and described in the form land parcels. If needed, the boundaries of the land parcels can be shown locally with legal effect. The real estate cadastre also documents the results of the soil assessment.

The cadastre in Germany performs the basic function for other areas. It should meet the requirements of legal relations, the administration, and the economy and, in particular, take appropriate account of the needs of the state planning, the development planning, real estate regulations, the determination of land plot values, and environmental and nature protection.

Today, the cadastre fulfills all legal demands and demands of administration and the private sector. It is a basic Land Information System (LIS) of great variety and flexibility in planning, environmental protection, etc. As mentioned earlier the cadastre shows the de facto status of land properties in both graphic and textual records. Today, maps and cadastral records are stored in computer systems throughout Germany. Although cadastre in Germany is the responsibility of the 16 states, the computerized systems are unique with some small exceptions. Since the early 1980s, cadastral computer systems have been developed and implemented step by step. These systems are the automated cadastral map and the automated property register.

The former parcel register was operated in a digital system called the **Automated Property Register (ALB)** in most parts of Germany. Cadastral maps were fully digitized in Germany. This target system was called the **Automated Cadastral Map (ALK)**. Both systems were more or less harmonized in all states, so that a nationwide user could have access to the same structure of data across the whole nation. Both information systems, ALB and ALK defined the basic Land Information System. However, both systems were based on technology of the last century, and the maintenance effort increased year by year. So, German mapping and cadastre authorities decided to develop a modern system based on state-of-the art technology. Today, this new system is called the **Automated Cadstral Information System (ALKIS)** and has replaced ALK and ALB. ALKIS stores all information in one object-oriented database system in compliance with the standards set by the ISO/TC 211 and the Open Geospatial Consortium.

20.3.1 Content of ALKIS (Nonspatial Information)

ALKIS contains the following descriptive information for a parcel:

- Cadastral district
- Parcel number
- Parcel order
- Legal status
- Location
- Legal area of a parcel and parcel segment
- Results of land valuation (soil validation)
- Type of land use
- Building descriptions
- Other descriptions
- Remarks
- Internal information about year of creation of the parcel, year of maintenance, number of cadastral map, and number of survey plans.

The content of ALKIS can be provided in different data exchange formats as well as user-friendly standard products harmonized in all German states. Figure 20.7 shows an example for a legal description of a property.

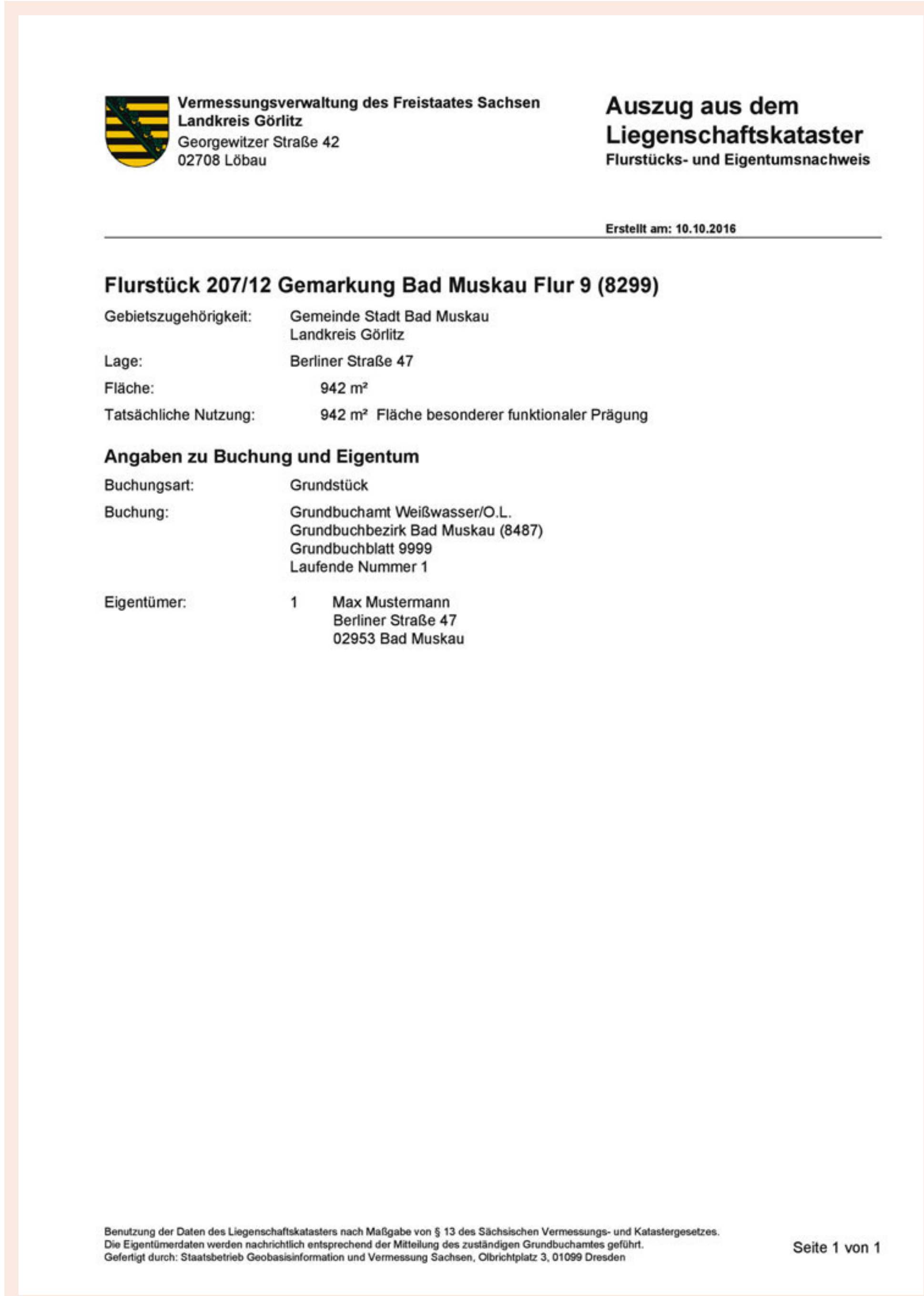
ALKIS therefore textually describes all parcels of the area of the national territory. All information concerning boundaries, geometric extension, land use, and location of parcels are determined in cadastral maps and cadastral registers. However, the land register keeps information about ownership, leasehold, and other legal appointments. So, the land register and the property cadastre depend on each other. An exchange of information is urgently needed and implemented. Therefore, an automated data transfer process was established in order to keep consistency between both systems. The data transfer in Bavaria, for example, is specified as follows:

- The technical register data and owner register of the ALB is redundantly implemented at the land register.
- The land register at the local court and cadastral office have a complete ALB database at their own disposal.
- The land register maintains the legal entries and the owner information in their own database.
- The cadastral office does the same with the information related to the parcel.
- All changes are fully documented.
- Data are automatically updated each night (differential data update) by a network connection.
- After each transmission, the data are identical in both databases.
- The cadastral office is allowed to change addresses as well.
- This data transfer process has been operational since 1996 in all Bavarian cadastral offices (79) and land registers (104).
- The transfer of paper documents is no longer necessary.
- Each year, both databases are checked in terms of consistency.

This data exchange process is still in use and will be replaced by a modern solution based on web services once the land register offices have implemented an object-oriented land register.

20.3.2 Content of ALKIS (Spatial Information)

In former times, nondigital cadastral maps generally existed in the form of grid maps scaled at 1 : 1000 or at 1 : 5000 based on Gauß–Krüger coordinates [4]. In Germany, all analogue maps have been replaced by digital maps (ALK) and databases containing geographical information about:



Vermessungsverwaltung des Freistaates Sachsen
Landkreis Görlitz
Georgewitzer Straße 42
02708 Löbau

Auszug aus dem Liegenschaftskataster
Flurstücks- und Eigentumsnachweis

Erstellt am: 10.10.2016

Flurstück 207/12 Gemarkung Bad Muskau Flur 9 (8299)

Gebietszugehörigkeit: Gemeinde Stadt Bad Muskau
Landkreis Görlitz

Lage: Berliner Straße 47

Fläche: 942 m²

Tatsächliche Nutzung: 942 m² Fläche besonderer funktionaler Prägung

Angaben zu Buchung und Eigentum

Buchungsart: Grundstück

Buchung: Grundbuchamt Weißwasser/O.L.
Grundbuchbezirk Bad Muskau (8487)
Grundbuchblatt 9999
Laufende Nummer 1

Eigentümer: 1 Max Mustermann
Berliner Straße 47
02953 Bad Muskau

Benutzung der Daten des Liegenschaftskatasters nach Maßgabe von § 13 des Sächsischen Vermessungs- und Katastergesetzes.
Die Eigentümerdaten werden nachrichtlich entsprechend der Mitteilung des zuständigen Grundbuchamtes geführt.
Gefertigt durch: Staatsbetrieb Geobasisinformation und Vermessung Sachsen, Olbrichtplatz 3, 01099 Dresden

Seite 1 von 1

Fig. 20.7 Example of an ALKIS standard product (property information)

- Parcel boundaries and boundary points
- Parcel number
- Boundaries of cadastral and municipality districts
- Survey and geodetic control points
- Buildings
- Building numbers
- Street names
- Results from official soil assessment
- Type of land use
- Basic topographical details like pavement curbs, bicycle tracks, trees, dams, walls, etc.

Also for graphical information ALKIS provides standard products such as cadastral maps as shown in Fig. 20.8. Cadastral maps can be produced in color or as monochrome pictures.

Since the content of ALKIS is stored in digital form, extracts can be composed individually depending on the re-

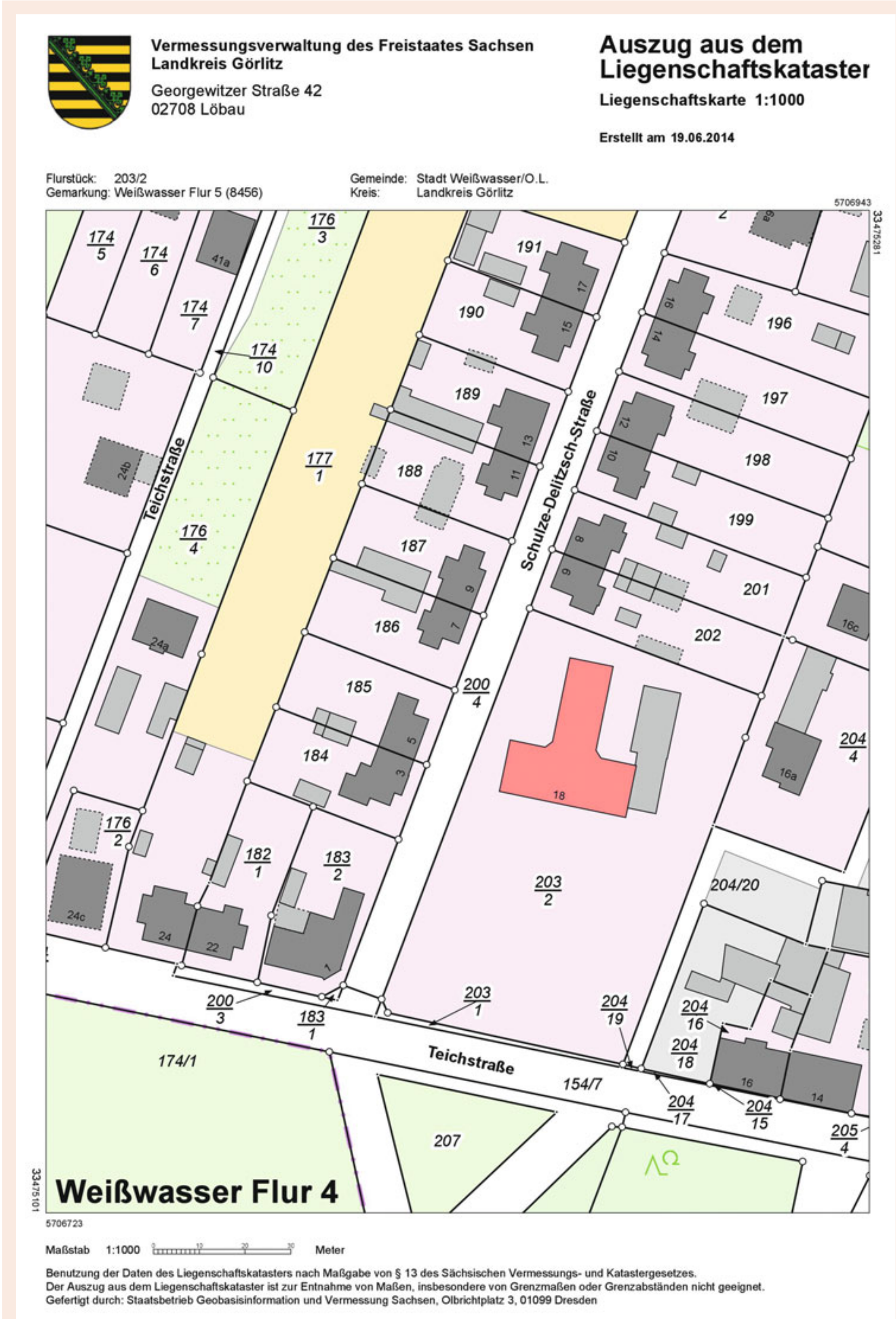


Fig. 20.8 Example of a digital cadastral map

quests of the user, for example as analogue listings, plots, or digitally in different, well-defined formats (e.g., XML).

Cadastral maps must be feasible as a basis for development plans and for the revision of official topographic map series. According to the legal tasks of the cadastral maps the contents is based mainly on terrestrial surveys (boundaries, buildings). Topographical details may sometimes be put into the maps by photogrammetrical methods. In areas where the cadastre has a high geometric accuracy, it is possible to create new legal boundaries from existing plans by calculating without surveying in the field.

20.3.3 The Integrated Cadastre Information System in Germany

The developments of the components of the cadastral databases for maps (ALK) and textual descriptions (ALB) started the 1970s and 1980s. They were developed by the surveying and mapping authorities in the technical environment and possibilities of that time. Thematic links between ALK and the topographical information system ATKIS are very difficult to establish because of different definitions in the feature catalogues. The models did not follow national or international GI standards.

The Working Committee of the Surveying Authorities of the States of the Federal Republic of Germany (AdV, Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland) decided to design a new and future-oriented system ALKIS in combination with a redesign of the Official Topographic and Cartographic Information System ATKIS. It should also consider the geodetic reference points of the Automated Reference Point Information System (AFIS). So, the new integrated system is also called the AAA data model [5] (for further information, see section titled "AAA Data Model – the Step to Standardized Official Reference Data").

Within the AAA data model the combination of ALKIS with ATKIS is designed to:

- Process all necessary cadastral and topographical data for a parcel based map and register of landowners, land use, and more unified basic data for the entire country.
- Control the use and maintenance of the system.
- Enable the use of the entire geographical data of the surveying authorities for all users via a metadata system including quality information for all data and a standardized data interface for ALKIS and ATKIS. Of course, existing links of user-specific data to the ALK, the ALB, or the ATKIS are transferred to the new systems without major additional costs.

The modeling of the AAA application schema is fully based on the ISO standard Unified Modeling Language (UML) and with respect to the definition of the data interface on the Extensible Markup Language (XML).

AAA Data Model – the Step to Standardized Official Reference Data

As discussed above, a number of independent geographic information systems have been developed in the field of cadastres and topographical mapping in the last three decades. The new approach, called ALKIS (Official Real Estate Cadastre Information System), was launched to harmonize the models of ALK and the topographic database ATKIS and to integrate the cadastral map and the land titles into one single model, which was separated into ALK and ALB in the past for historical and technical reasons (see the beginning of Sect. 20.3). Also adding the geodetic reference points (AFIS) almost all official data of the surveying and mapping agencies are now defined in a common and harmonized data model, called the AAA data model (AFIS-ALKIS-ATKIS data model).

For this reason, the **AdV** has started developing a new conceptual data model based on international GIS Standards, which helps fulfill this harmonization. This will provide the surveying and mapping agencies in Germany with a nationwide well-defined database that can be used as a baseline for many other thematic application schemas. For the increasing efforts in building up a spatial data infrastructure, the AAA data model can generally be used for the standardization of the related thematic data as well and help to standardize the geographic information in Germany brick by brick.

While the data model is a common approach for all states in Germany, the corresponding transposition of this concept and the software development as well as implementation are the responsibility of the states.

The task of the surveying, mapping, and cadastral authorities of the Federal States of Germany is to provide basic data for spatial referencing (Geobasis data) for the use in official, industrial, and private domains. The demand for this digital data is still increasing and was reflected very early on by the authorities, who set up the ALK (Automated Cadastral Map) and ALB (Automated Property Register), as well as the topographic data in **ATKIS (Official Topographic Cartographic Information System)** in a digital, standardized way in all German states. Most of the federal states of Germany are encouraged by a political or legal advice to use official reference data of the AAA data model as a basis for all other official thematic information systems.

In addition to AFIS, ALKIS, and ATKIS other extensive digital database inventories have also been introduced according to specific concepts, e.g., digital orthophotos, raster data of topographical state maps, and digital elevation models.

By integrating the former ALK and ALB databases in **ALKIS** (Official Real Estate Cadastre Information System) a lot of redundant information (e.g., topographic objects, buildings, land use objects), sometimes with different semantics, have been removed. So, a core objective was to harmonize the different semantics on the level of the conceptual data model. Additionally, there has been a great deal of inconsistencies in the separately driven databases of ALK and ALB. With the migration to ALKIS these inconsistencies have been removed and will not occur again.

Digital Terrain Models (DTM) are no longer a specific object group within the digital landscape model of ATKIS but are now defined as a separate application schema. Similarly to geodetic control point objects, the universality of DTM as an independent database is now given and, therefore, also the opportunity to create new combined data inventories or products using data from any application system.

In Germany, geoinformation of official surveying and mapping also includes information about the geodetic control points. Since these data originally belonged neither to ALK nor to ATKIS, they are now modeled in their own information system called the **Official Geodetic Control Station Information System** (AFIS) with a separate feature catalogue.

The AdV projects AFIS, ALKIS and ATKIS, with their nationally standardized features are described in a common form under the heading *Documentation for Modeling Geoinformation of Official Surveying and Mapping*. They are associated with each other in a **common reference model** as a **common application schema for AFIS, ALKIS and ATKIS** (see next sub clause).

The common application schema provides for the recording and management of **metadata and quality data** in accordance with ISO standards.

The former ALK system was originally used in surveying organizations, performing as a central information system to fulfill all tasks that had to be carried out related to cadastral issues. Moreover, from the very beginning an objective of the ALK was also to provide basic geographic data for many GIS applications in various fields in local governments, utilities companies, and for other customers. Right now, there is also a great demand for access to these valuable data from other areas, like the financial sector, lawyers, and notaries. Generally, all users need up-to-date data for their applications. So, a fast data transfer or even online access to these data using SDI technology could be very helpful in future. Currently, many projects are established to build up a national spatial data infrastructure. This will help to enhance the access to public geodata and reach interoperability between different geographic information systems within various state administrations.

Therefore, the approaches aiming at GIS interoperability, like those based on the concepts of ISO (International Organization for Standardization) and OGC (Open Geospatial Consortium), are very useful in this field and have to be taken into account by modeling the new AAA application schema. The goal of the international standardization is to create foundations for the common, holistic, and cross-domain use of geodata at various locations by individuals, applications, and systems based on a standard description of the content of existing or planned data inventories, the functionalities of data processing and communication. The modeling is based on the results of ISO/TC 211 in the form of the 19100 series of standards at their current stage of processing. The data exchange interface also uses parts of the OGC specifications.

The AAA application schema also provides concepts for the following issues:

- A data versioning concept is integrated for historicizing all object classes and to allow a differential updating of secondary databases.
- With the semantic harmonization of cadastral and topographical information, the conceptual possibility is given to store all the spatial data of the cadastral and mapping agencies in just one database. Practical implementations that are able to solve the challenge of generalization are currently research issues, but the theoretical possibility is given.
- The new data model provides clear advice as to how the data have to be collected and structured. Therefore, only fully compliant data can be stored in the databases. For example, the topological elements for cadastral information (surface, boundary, point) will secure consistent data by avoiding overlaps and gaps. It is expected that the quality of cadastral and other information will be significantly enhanced.
- 3-D application schema can also be modeled based on the AAA basic schema, which contains fundamental geometrical and topological elements.
- By the consequent use of ISO standards the AAA application schema is fully compliant with the INSPIRE data specifications and the LADM data model, which use the same model-driven approach.
- The data of the surveying and mapping agencies in Germany are described in a formal and transparent way using the conceptual schema language UML.

The concept of the AAA data model is publicly available at the website of the AdV, and the maintenance is guaranteed for the next years. This conceptual data model describes geographic and nongeographic features as well as their relations (associations). To describe this model in a standardized way it has been based on the ISO draft standards in the field of geographic information. The AdV decided to use a standardized conceptual schema language, the **Unified Modeling Language (UML)** to describe the application schema and the feature catalogue. This language is also used by ISO/TC 211 in the field of standardization of geographic information.

UML was developed by the Object Management Group (OMG) for the purpose of describing application schemas. In order to guarantee the standard use of UML in the 19100 family of standards, their application is specified in [6]. The purpose is the complete and unambiguously interpretable, formal description of the content and structure of data inventories. The description is independent of the type of implementation and the applied programming language. A standard description of all geodata can be achieved only with formal description languages. The UML application schemas can be automatically interpreted by suitable programs and translated into internal data and/or database structures.

The AFIS-ALKIS-ATKIS Application Schema

The application schema provides the formal description for data structures and data content in one or several applications. It contains the complete description of a database and, in addition to geographical data, may also contain other associated data. The fundamental concept of abstracting the real world means the introduction of thematic objects and of rules and regulations on how it is documented and managed. Thematic objects are classified by types. At the type level, the application schema describes the feature types of the real world. Data themselves exist at the instance level. They represent individual examples of a feature type in the real world and can be interpreted by the application schema, see also ISO 19101 *Reference model – Part 1: Fundamentals* and ISO 19109 *Rules for application schema* (Fig. 20.9; [7, 8]).

The purpose of an application schema is to achieve a common and unified understanding of data and document the data content for a specific application environment so as to obtain unique information about these data.

The common AFIS-ALKIS-ATKIS application schema offers a unified and object-oriented basic model for AFIS, ALKIS, and ATKIS, which wherever possible is to be depicted and managed using the commercially available GIS software. An application schema can use specifications from various subschemas. In the case of the AFIS-ALKIS-ATKIS application schema, mainly subschemas from the ISO 19100 series of standards are used (Fig. 20.10). In those areas, where there are as yet no ISO-standards, additional schemata of the Open Geospatial Consortium are used (e.g., Web Feature Service, Filter Encoding). That OGC specification will be replaced by the corresponding ISO standard once it is available. The implemented ISO standards are shown in Fig. 20.11.

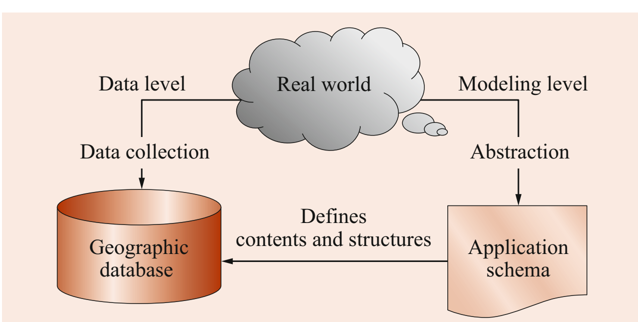


Fig. 20.9 Data modeling and application schema

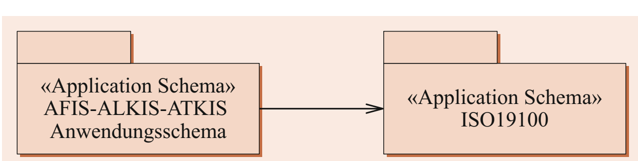


Fig. 20.10 Dependency of the AFIS-ALKIS-ATKIS application schema on the structures standardized from ISO 19100

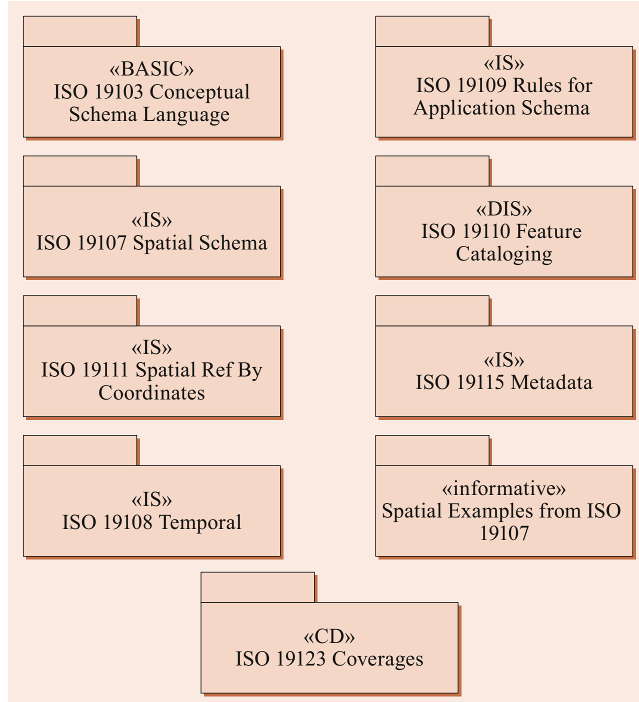


Fig. 20.11 Components used from the ISO 19100 series of standards

The AFIS-ALKIS-ATKIS application schema is subdivided into the **basic schema**, the versioning schema, and the **AFIS-ALKIS-ATKIS thematic or technical schema**. The basic schema is the basis on which thematic objects are modeled in the thematic schemas (Fig. 20.12).

The versioning schema shows the concept for historicizing thematic objects. An internal implementation schema is not part of common modeling. It is created by depicting a conceptual application schema in specific GIS systems as part of the implementation process. The application schema is the basis on which operations for data exchange and technical stipulations for data outputs are defined. Other standard applications schema are also integrated, e.g., web feature capabilities, filter encoding capabilities, etc.

Two important components are discussed in more in detail in the following: the AFIS-ALKIS-ATKIS Basic Schema and the new standardized data exchange interface NAS.

The AFIS-ALKIS-ATKIS Basic Schema

The following systematic is used for unique designation of the defined classes (Fig. 20.13):

1. Standardized classes maintain the standardized prefix in the class name (e.g., FC for "Feature Catalogue", MD for "Metadata").
2. Classes such as AFIS-ALKIS-ATKIS-specific additions to the standardized Feature Catalogue get the prefix AC.

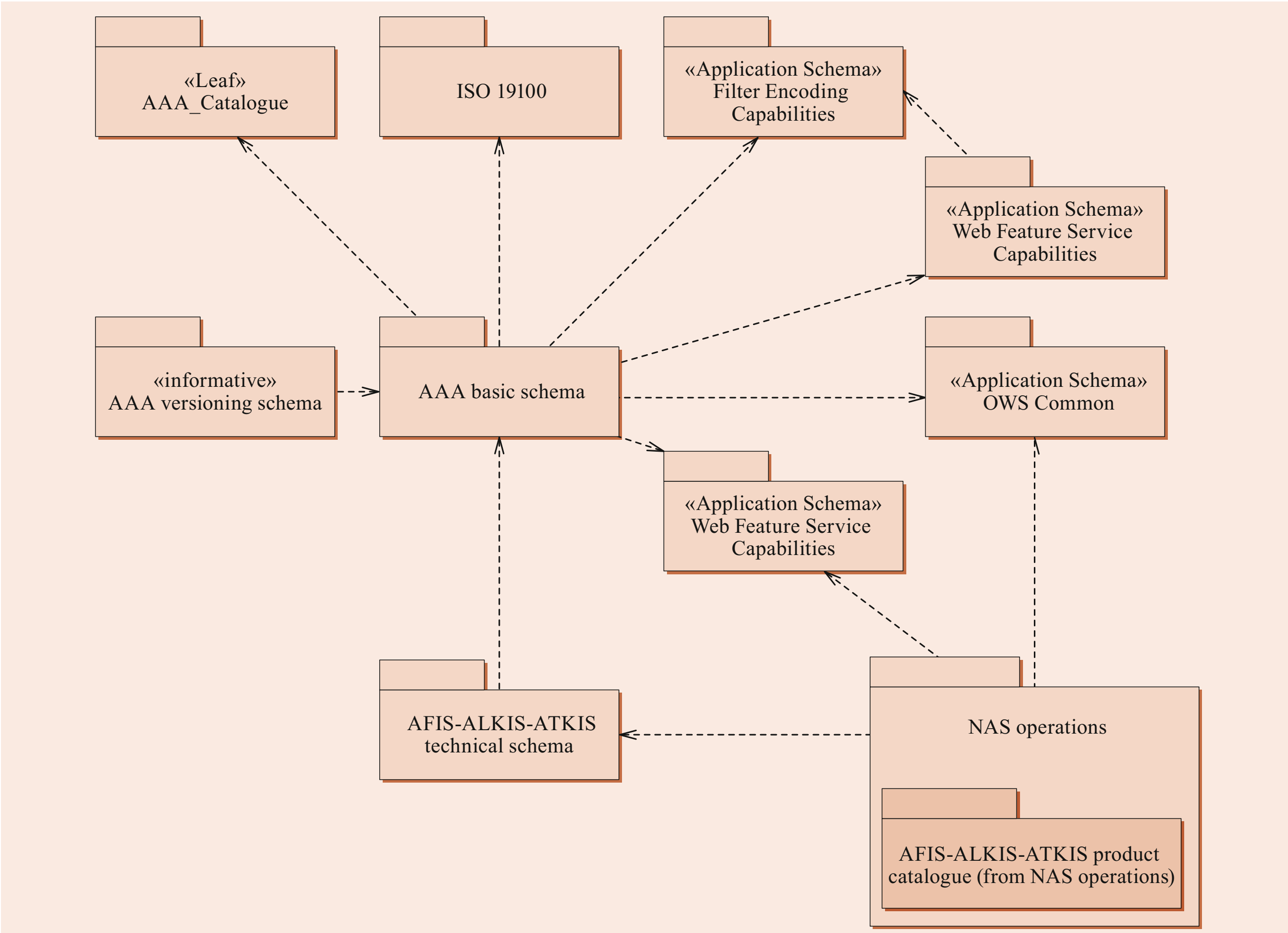


Fig. 20.12 The components of the AFIS-ALKIS-ATKIS application schema

3. Classes with fundamental meaning for AFIS, ALKIS and ATKIS get the prefix AA.
4. Classes derived from the ISO TS_*Component classes ("simple topology"), get the prefix TA; also the analogously created class for topological surfaces with multiple spatially separated geometries (TA_MultiSurfaceComponent).
5. Classes with commonly used geometries get the prefix AG.
6. Classes of independent geometries get the prefix AU.
7. Classes of presentation objects get the prefix AP.
8. Classes for the modeling of PointCoverages get the prefix AD.

The basic schema mainly contains elements that have been inherited from the relevant ISO standard and can be considered as a profile of the ISO standards. Only those elements are referenced that are needed for the data modeling of official geographic information systems in the public administration. These ISO elements have been completed by additional elements that can also be used in different thematic domains (e.g., presentation objects, process management, etc.). All thematic objects (e.g., the cadastral parcel) are modeled by inheriting these basic elements.

The ISO standards contain conceptual data models in a very abstract way in order to allow the application in various domains. Therefore, they must be specialized for the specific application schemas. Using ISO 19107:2019 *Geographic information – Spatial schema* as an example, the profile, as well as the necessary adjustments, will be discussed [9].

ISO standard 19107:2019 [9] provides spatially referenced basic constructs for use in application schemas; of these, solely the following constructs are used for AFIS, ALKIS, and ATKIS, in order to reduce complexity (Table 20.1).

The geometric and topological objects are described as UML classes. The standard also contains spatial operations, which the geometric and topological objects (GM_Object or TP_Object) use as parameters (create, delete, change, spatial evaluations, ...). The defined classes have no direct application, i.e., they are not instantiable. Their use in special application schemas is achieved through inheritance; inso-

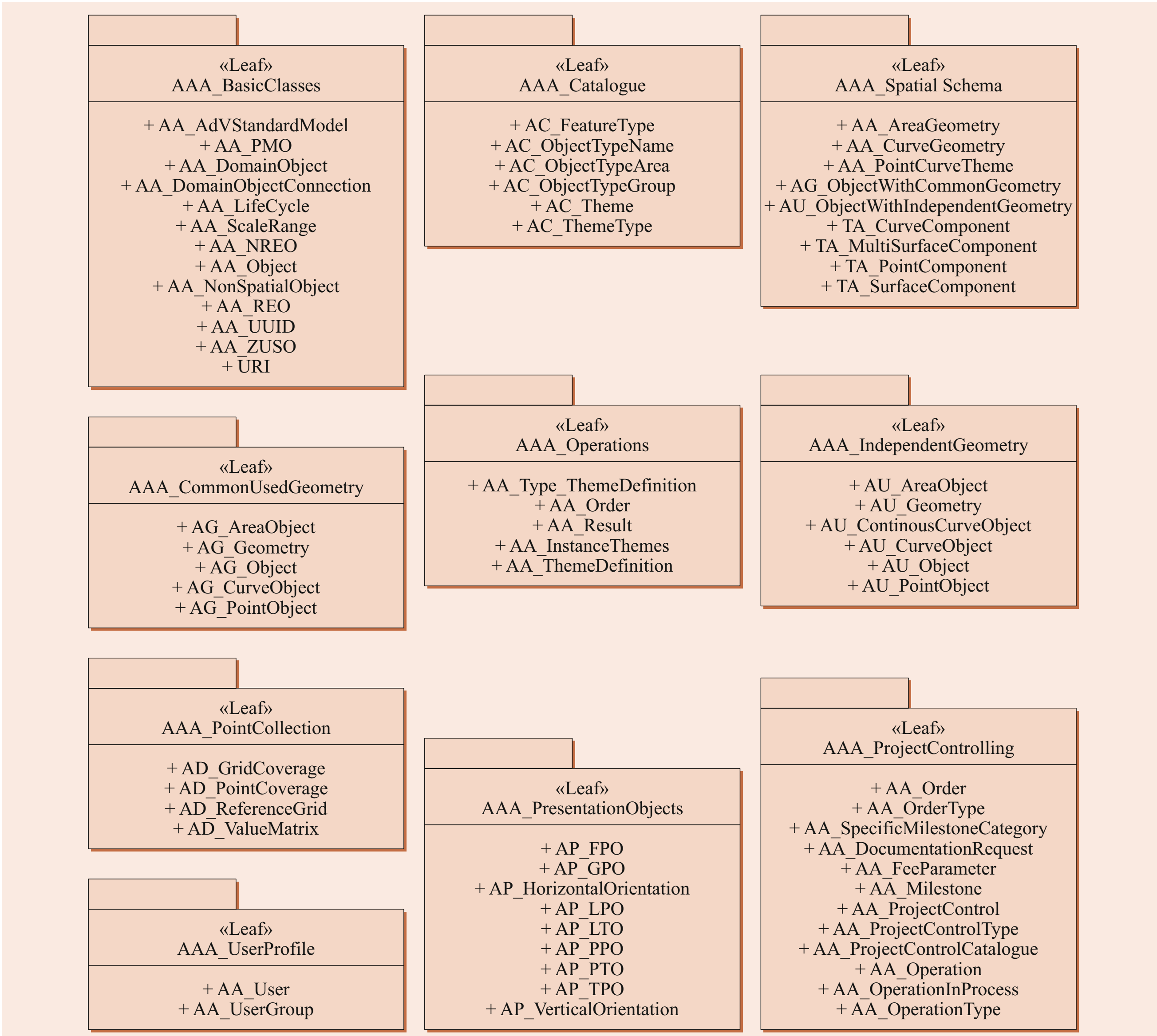


Fig. 20.13 Main packages of the basic schema

far as the classes of the Spatial Schema for AFIS, ALKIS, and ATKIS are not supplemented by special attributes, they are directly used; however, in this application they are used for simplification purposes.

The inheritance of the ISO elements is shown in Fig. 20.14. The applied ISO elements carrying the prefix "GM" are modeled as attributes within the AAA feature types (prefix "AA"). These feature types will be used in thematic feature types by inheriting all elements into these classes. For example, the class AA_CurveGeometry can inherit all its elements to a thematic class "ParcelBoundary" defined in the application schema. Additional requirements and constraints are described in specific notes linked to the corresponding elements.

The geometric primitives usually appear as attributes of objects; this does not mean, however, that the geometry is always redundant in principle. The common AFIS-ALKIS-ATKIS application schema has the following options for linking the spatial reference:

- The formation of node-shaped, edge-shaped, and face-shaped objects with "simple topology". Additionally, face-shaped objects with "simple topology", which consist of two or several spatially separated faces (required for modeling of land parcels with multiple, nonadjacent parts). The ISO schema "Simple Topology" is used, which expresses topological features by geometrical features, while still offering topological functionality.

Table 20.1 ISO 19107 profile for the AAA application schema [9]

Geometric objects (GM_Object)			Topological objects (TP_Object)	
Geometric primitives	Geometric complexes	Geometric aggregates	Topological primitives	Topological complexes
GM_Point GM_Curve GM_PolyhedralSurface	GM_CompositeCurve GM_CompositeSurface	GM_MultiPoint GM_MultiCurve GM_MultiSurface	TS_PointComponent TS_CurveComponent TS_SurfaceComponent TS_Face	TP_Complex

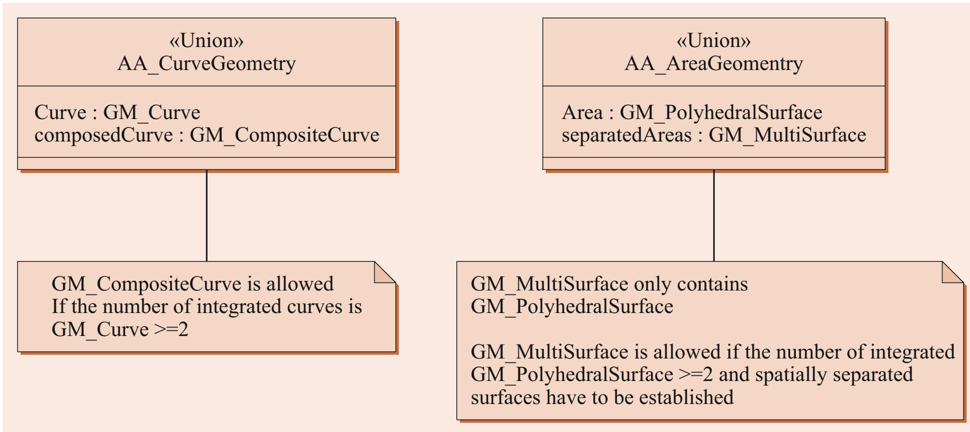


Fig. 20.14 Necessary enhancements and constraints of ISO 19107 *Spatial Schema* [9]

- Formation of point, line, surface, and volume objects, which share lines and points.
- Formation of point, line, surface, and volume objects with "independent" geometry.
- Formation of topological and geometrical "topics" that allow the selective combination of feature types into complexes, in order to express geometric identities and/or topological correlations.

Each spatially referenced AFIS-ALKIS-ATKIS feature (AA_REO) refers to a maximum of one geometry. Should it be necessary to keep several geometries for a real-world object (e.g., generalization, various coordinate reference systems, point and surface geometry), a separate feature (where necessary as a map geometry object) should be formed in each case.

All necessary enhancements and restrictions of the ISO Spatial Schema are summarized in the AAA application schema.

Standard-Based Data Exchange Interface (NAS)

The standards-based data exchange interface (NAS) is used when it is necessary to exchange geographic information that has been modeled on the basis of the common AFIS-ALKIS-ATKIS application schema. This can relate both to information that has the same structure as the stored data inventories, including the additional data (presentation objects, map geometry objects), and also to information from derived views on these data inventories (e.g., output feature types) but not to data inventories for which the object reference is completely lost (e.g., purely graphically structured data) or data that is to be defined according to a different basic schema (e.g., DXF data).

Accordingly, NAS is used wherever the application emphasis is on the following, in line with the user's requirements,

- The originality of the data
- The full evaluation capability
- The feature-specific updating.

The NAS can be automatically derived from the AAA application schema and is based on international GIS standards (ISO 19118 [10] and OGC GML 3.0 [11]).

The standards AFIS, ALKIS, and ATKIS of the AdV are described in this document in conceptual format on the basis of ISO 19109 Rules for Application Schema [8]. This means specifically:

- Modeling in UML with the software tool Enterprise Architect
- Compliance with the regulations of ISO/TS 19103 [6] for the use of UML
- Use of ISO 19107 (and, therefore, by implication ISO 19111 [4]), ISO 19115, ISO 19123
- Automated derivation and mapping of feature catalogues in accordance with ISO 19110.

An automated derivation of the interface for exchange of AFIS, AKIS, and ATKIS objects, NAS, completes this picture (Fig. 20.15).

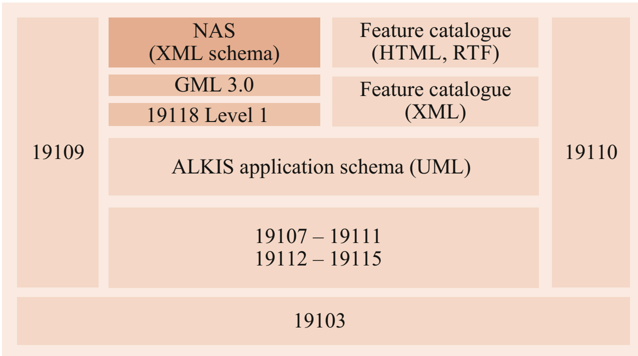


Fig. 20.15 Derivation of the NAS encoding rules and the embedded ISO standards

For this purpose, ISO 19118 Encoding [10] defines a framework document for the creation of what are referred to as Encoding Rules, to derive interface definitions for data exchange from a UML application schema. The de-jure standard also describes in an informative appendix, special Encoding Rules for the creation of XML schema definitions. The variability permitted for mapping UML to XML schemas does, however, result in the ability of ISO 19100 basic norms to be converted in different ways. A stipulation by the AdV would result in AdV-specific interfaces, and, thus, the purpose of the de-jure standards, namely, to achieve inter-operability, would not be fulfilled. There are currently no standardized XML schemas for the basic ISO-standards.

Besides the official de-jure standards for defining interface definitions, there is a de-facto standard, based on the Geography Markup Language (GML) [11] of the Open Geospatial Consortium (OGC), for encoding geoinformation, which is also provided for standardization in the ISO 19100 series as ISO 19136. GML is a system for rules in modeling user-specific objects with their specific features in XML schema. For the description of objects and their numerical, textural, geometrical, temporal, and other features, a number of standardized components (e.g., geometry types) can be used. This way, a user-specific XML is defined in a GML-application schema. GML is gradually achieving market acceptance.

Besides coding of thematic objects, NAS contains operations for communication with a data storage system (updating, inserting, locking/unlocking objects, reservation, request for output products, user-specific updating of secondary database (NBA)) modeled in the application schema; GML-features are embedded in corresponding, principle web-service-qualified operations types using elements of the Web Feature Service (WFS) and Filter Encoding Standard (FES) of OGC. In this sense, an AFIS-ALKIS-ATKIS-data storage can be compared with a capsulated Web Feature Server, which considers additional AFIS-ALKIS-ATKIS-specific demands.

The AdV uses the new development of AFIS, ALKIS, and ATKIS to pursue the objective of creating the basis for a common, unified, and interdisciplinary use of geodata. In this sense, existing or foreseeable standard functionalities of application software should be used wherever possible. One example is the NAS described in this chapter. The use of AdV-specific solutions is largely avoided. Due to the current status of international standardization in the field of metadata and the operations for updating and requesting of GML-data, however, this is currently only possible to a certain degree and with the AFIS-ALKIS-ATKIS specific add-on.

For these reasons, the AdV has decided on the following procedure:

- The framework document for Encoding Rules defined in ISO 19118 [10], Sect. 8 is applied for the NAS (Level-1 – Conformity with ISO 19118).
- "NAS Encoding Rules" according to ISO 19118, Section 8 are defined and documented. These "NAS Encoding Rules" map the conceptual AFIS-ALKIS-ATKIS technical and basic schema on a GML-3.0 application schema. This schema file is specified by the schema AAA.Fachschema.xsd (including additional schema files).
- The conceptual schema of NAS-operations is set up by using "NAS Encoding Rules" and WFS/FES-xml-schema in XML schema components and specified in the schema file NAS-Operationen.xsd.
- The "NAS Encoding Rules" are stipulated in conformity with GML 3.0 and are, furthermore, as simple as possible and as far as possible in conformity with ISO 19118 "XML Encoding Rules". This means that NAS uses GML in accordance with GML-specifications.
- An automatic derivation of NAS is supported, which means that the XML schema definitions of NAS can be derived using the "NAS Encoding Rules", the UML application schema, and additional regulations formally described in the form of control parameters.

It is important to note that the stable conceptual model is fully described in the UML application schema. Future adaptations to the IT/GI Mainstream will also become necessary for depiction on specific implementation models (e.g., XML representations).

The AAA Application Schema as a Brick of a Spatial Data Infrastructure

The methodology for setting up a spatial data infrastructure is, in principle, independent from the geographical extent that will be covered. Basically, a spatial data infrastructure should meet the following requirements at least:

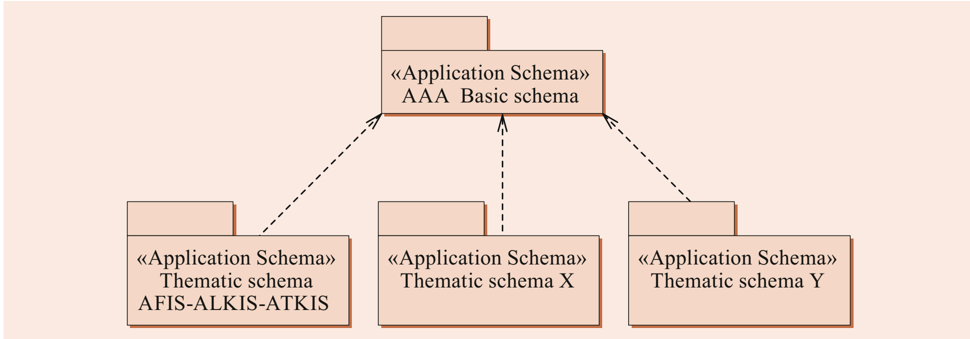


Fig. 20.16 The basic schema as a basis for the modeling of application-specific thematic schemas

- Implementation of controlling mechanisms for compliance to the defined SDI standards (monitoring and reporting)
- Organizational framework for the creation process (who will make the decisions?)
- Defining conditions for data access and data use
- Collecting and providing metadata for discovery services
- Providing the data and services within a geoportal
- Defining harmonized geodetic reference systems
- Harmonized data and services
- Standardized description of the provided data (model-driven architecture) and based on these methods for model transformation.

Most of these points are also addressed by the **INSPIRE directive** for creating a spatial data infrastructure in Europe. For the last two items, the AAA data model provides methodology that, in principle, can be used for official surveying and mapping data, as well as in other domains.

With the legal implementation of the SDI coordination group Germany has reflected the requirements above and decided on a three-step approach for implementing the SDI in Germany

- Step 1. Development of data access based on the ISO 19115 metadata standard and the corresponding catalogue services for queries
- Step 2. Model harmonization using GI standards (possibly also the AAA model) and decision on commonly used SDI-standards or profiles (e.g., WMS, WFS, Catalogue Service – CSW, . . .)
- Step 3. Development of a geoportal.

In principle, no one will be forced to implement the AAA data model in any other domain. On the other hand, the AAA application schema provides a framework and methodology that could be widely used to describe and provide the spatial data in a standardized way. This will open possibilities to map between one data model to another, maybe to handle semantic model transformation. This approach will significantly enhance interoperability within a SDI.

How to Use the AAA Application Schema in Other Thematic Domains

The AFIS-ALKIS-ATKIS basic schema (AAA basic schema) forms the basis for the technical modeling of AFIS, ALKIS, and ATKIS objects and for the data exchange process. The thematic schemas are created from this basis. Its application is not limited to just AFIS, ALKIS, and ATKIS. Other technical information systems can also use the classes defined in the basis schema to model their thematic schema (Fig. 20.16).

The basic schema is subdivided into ten packages (see above) containing fundamental elements that are necessary for a data model (e.g., a unique object identifier). These packages are independent of any thematic content and can, therefore, be used completely or partially, depending on the requirements of the thematic application. Generally, it is also possible to use not just feature classes of the basic schema but also appropriate thematic feature classes, like "building" or "parcel" by adding specific properties or attributes.

Example: Using a Parcel for Municipality Issues

In the case of integrated modeling of municipality content, it could be necessary to describe additional information for the existing parcel feature provided by the surveying authorities. A possible approach could be the following modeling example (Fig. 20.17).

The prefix enables us to distinguish between feature classes of different domains, although they are defined within one data model. "AX_" means the AAA data model, and "KS_" means the specific information model of a municipality. The required information of whether the real estate is owned by the municipality is modeled as a derived attribute, because this information is already available in the AAA data model. By modeling adiitional information as an extension to the AAA data model redundencies in the source data will be avoided.

The benefit of using the same methodology, or even the same feature classes, is to be able to reach a common understanding of the provided data. If you want to provide not just maps but also object information, it is necessary to have a standardized description of the data content. The standard

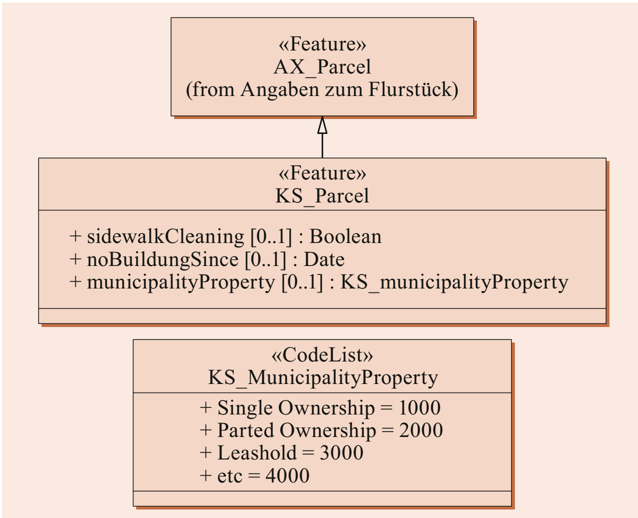


Fig. 20.17 Integrated modeling of different thematic domains

series of ISO/TC 211 (ISO 191xx) provides a framework for this. Additional web service (WFS, Web Feature Service) are able to deliver these object information in a standardized way.

The official geospatial data are already widely used. By integration of various datasets in the AAA data model or using the same methodology the use and further processing of public geospatial data will be simplified substantially. The main reason is the consistent application of the ISO conceptual standards in the field of geographic information and using web services (WMS, WFS) to provide the data. So, some essential advantages will arise for users by the new conception:

- The implementation of the concept under consideration of international standards will ensure investment safeguarding, vendor independence, and standardization of public geospatial data.
- A universal, browser readable interface (XML encoding) for all public geospatial data.
- The AAA basic schema becomes a core data model that can easily be combined or extended with other data from various administrations in order to build up a spatial data infrastructure in Germany.

Multidimensional Cadastral Systems in Germany

These days, economy, science, and administration have an increasing demand for official three-dimensional spatial information (3-D geodata) as a base for multiple applications. The surveying and mapping administration in Germany has accepted this demand as a challenge to develop and realize sustainable conceptions for 3-D geodata, focusing on fast and economic solutions. In this context, national and international standards, infrastructures, and activities had to be considered.

In recent years, the information systems of surveying and mapping and cadastres have been focusing increasingly on demands for the fourth dimension, which means the time component. Looking back to historical situations is needed by several applications, such as environment protection, urban planning, disaster management, and cadastral inquiries during disputes at court. The fourth dimension (life-cycle information) is now an integral component of the new German cadastral information system.

The Demand for 3-D Building Information

In Germany, the government targets of climate and environmental protection are currently leading to extensive changes in the energy sector, the so-called energy turnaround. This includes the end of the use of nuclear energy by 2020, the reduction of greenhouse gases and other objectives. As a result, planning processes especially have to take into account the use of photovoltaic technology, geothermal energy, wind energy and the energetic isolation of buildings (Fig. 20.18).

From the process view, data must be available to provide actual information on the environment and all energetically relevant topics. Very often, this leads to a data collection or at least to a data processing task. With the required information, the analysis and evaluation will give a sustainable picture of the energy balance, including possible savings, the use of renewable energy, and thermal insulation of buildings.

Three-dimensional geometry and semantics, particularly of buildings, are also particularly important for simulating and mapping of noise expansion to develop noise protection measures in cities. With a European directive every 5 years, the member states of the European Union are obliged to determine and to document noise pollution in cities. In addition, the progress of noise reduction is checked (Fig. 20.19).

The use of cadastral information for urban planning was always essential in the 2-D world, especially to consider property distribution. Nowadays, 3-D information is a basic demand of the urban planning sector. Demographic effects and other restrictions could be visualized in planning alternatives (Fig. 20.20).

Increasingly, 3-D information is furthermore used in the simulation of disasters, for example for evacuation and flood scenarios.

Fig. 20.18 Photovoltaic map of the city of Düsseldorf

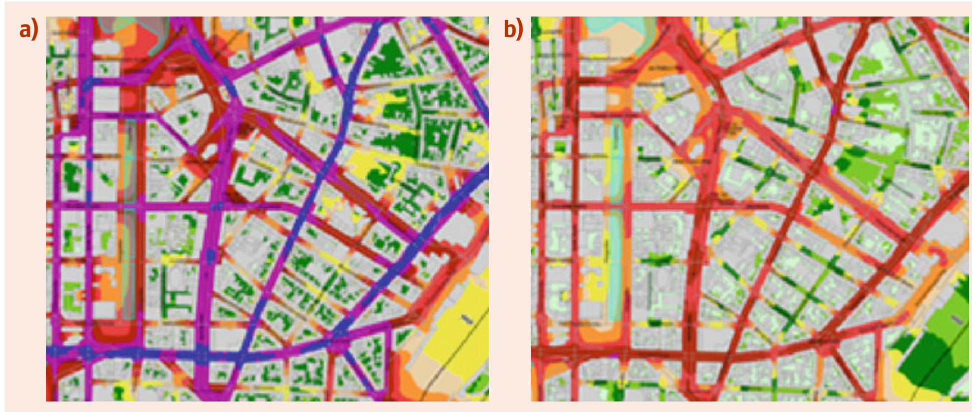


Fig. 20.19 Noise map of the city of Düsseldorf (**a** during the day, **b** at night)

Fig. 20.20 Planned school in the county of Recklinghausen

3-D Spatial Information in the AAA Data Model

The AAA data model is the only national standard for official spatial information in Germany. It was built up completely by the specialization of international standards. Using the model-driven approach, the data model can easily be extended by additional features, such as 3-D information for buildings. This was done recently for the AAA data model.

The integration of 3-D buildings takes into account the source of the data and the production process. The "legal" 2-D property building layer as major cadastral information is merged with the 3-D layer from a laser scan as a topographic source. The result is a "legal" 3-D building model.

For the German cadastre with over 50 million buildings, it is therefore of fundamental importance to store actual 3-D building information conforming to the AAA data model. The data model defines the official AdV product "3-D building model". As a consequence, the demand, especially on the economy, for official (administrative) 3-D building information can be fulfilled. In addition, this data participates in the existing national and international spatial data infrastructure (SDI), for example through simple export to the defined INSPIRE topics (Fig. 20.21).

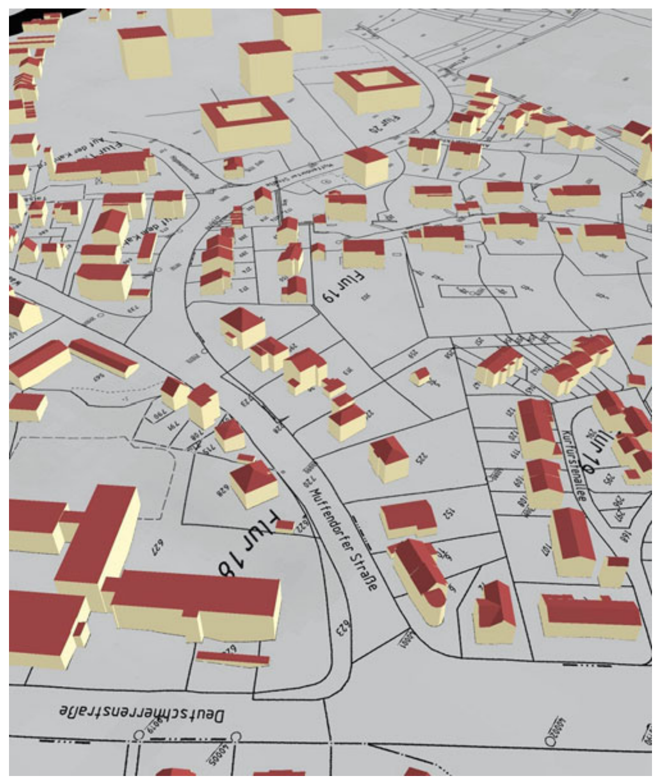


Fig. 20.21 View of a cadastral map with 3-D buildings (Bonn, Germany)

The Fourth Dimension (Time Component)

Traditionally, in the German cadastre, every parcel change (e.g., subdivision) is documented by surveying sketches and textual documentation. The development of the cadastral map is monitored continuously, and every change over time can be restored should the cadastre be disputed, usually using nondigital paper documents. Therefore, the modern possibilities for inquiries were also a technical requirement for the AAA data model. Besides this more internal cadastral use case, there are many further requirements for time-related cadastral information, such as:

- Monitoring the development of cities and villages over time
- Statistics of changes of land use and land cover
- Planning purposes
- Historical archiving
- Monitoring cultural heritage.

For each object, the AAA data model requires a unique identifier together with a designated time stamp for the creation and deletion of an object. However, once an object has to be deleted during an updating process, it will not be physically removed from the database. Only the life cycle of the thematic relevance has ended, not the existence of the object as an instance. A "deleted" object is then considered as historical information, which can easily be distinguished from the actual information. Sometimes, there are changes of an object that do not require the deletion of the object (e.g., only the name of a person changes). In that case, also the different versions of an object can be stored. Within the AAA data model this approach is, therefore, called a "versioning concept". Since every object carries life-cycle information, the storage of historical objects and versions of objects is not limited to any specific object type.

In the AAA data model, this approach is used to provide historical information as well as the incremental update of the secondary information systems used.

20.4 Cadastre and Land Register in The Netherlands

In The Netherlands, cadastre and the land register are organized into one central agency, The Netherlands' Cadastre, Land Registry and Mapping Agency (in short: Kadaster). Its origins lie in Napoleonic times, where it was originally set up as fiscal cadastre (the registration of properties for taxation purposes). National coverage of the cadastre was established in 1832, and since 1838 citing the cadastral land parcel number is obligatory in notarial deeds of transfer and deeds of mortgage. The cadastre thereby became a key to the land register. The fiscal cadastre also became a juridical or legal cadastre, a situation which is still a benefit today. A major revision of the Civil Code came into power in 1992. At the same time, the Land Registry Act was introduced, as a specific elaboration of the parts in the Civil Code pertaining to the system of property rights (to an immovable object) and its aspects of registration and cadastre. This constituted the land register and cadastral maps as a multipurpose system aimed at providing legal security of tenure, facilitating the land market, and supporting many government activities like spatial planning, development control, public acquisition of land, land taxation and management of natural resources.

In 2004, the Topographical Service of The Netherlands was integrated into the Kadaster organization (also making it the national mapping organization). Moreover, Kadaster has been assigned additional tasks in the field of digital government, primarily acting as the information node for spatial information. This way, Kadaster evolved into the organization it currently is, with (legal) responsibilities for the national reference network, cadastre, land register, and mapping, and as national information node for spatial information, ranging from addresses to cables and pipelines. Within the Kadaster organization, the cadastre and land registry tasks, however, have remained the largest activity.

As a public agency, Kadaster operates in a wider context within the National Spatial Data Infrastructure (NSDI), the system of key registers in The Netherlands and the Environmental Planning Act. Kadaster, therefore, cooperates closely with other public bodies (in particular municipalities), notaries, utilities, and other agencies, such as the National Geological Survey and Statistics Netherlands. In The Netherlands, the Ministry of the Interior is politically responsible for digital government and spatial information. More information can be found at www.kadaster.com.

20.4.1 System of Key Registers

In The Netherlands, several state and municipal registers have been designated by law as a key register [12]. The cadastral registration (holding both the administrative data and the cadastral map) is one of them. The key register status implies that specific data elements are labeled as authoritative and must be used by all public authorities for their legal tasks and processes. For example, data on parcels are only to be acquired from the key register cadastre. For data to be authoritative, this requires high levels of currency, quality, availability, and accessibility. One of the characteristics of key registers is that public bodies are required to report (apparent) errors. In its turn, Kadaster has to process this feedback and, if necessary, take corrective action. The key register cadastre itself is linked to the key registers for persons, businesses, addresses, and large-scale topography and feeds into the key registers for valuation and topography (Fig. 20.22). The key register cadastre is open access (the cadastral map is even open data) and is, in practice, a pivotal register in the information provision in the land market [13].

A data catalogue underlies the system of key registers. We have experienced that semantic interoperability between different themes (be it between data in general or specifically for spatial and cadastral data) is crucial (Fig. 20.23).

The system of key registers provides the data backbone for e-government in The Netherlands. The scheme clearly indicates that the key register cadastre is well integrated in the

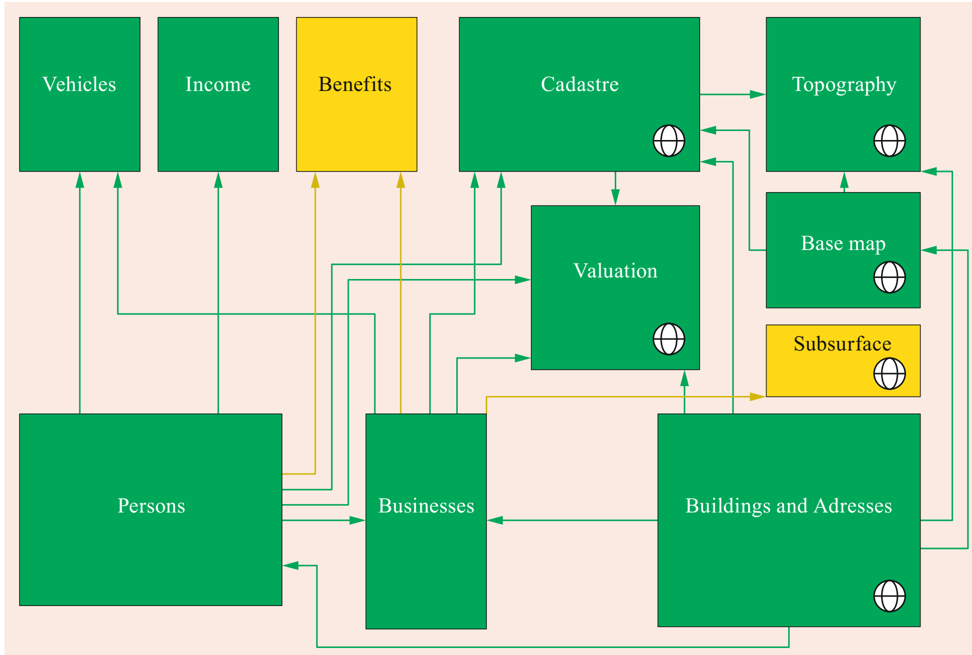


Fig. 20.22 System of key registers in The Netherlands. The registers in *green* are fully operational, and those in *yellow* are under development. The *arrows* indicate explicit coupling of objects; the small *globes* indicate where the implicit coupling by geometry is possible

Fig. 20.23 Excerpt from the catalogue of the system of key registers for the key register cadastre (www.stelselvanbasisregistraties.nl/registraties/)

national data infrastructure. The Dutch government pursues an active policy for data in terms of availability, accessibility, usability, and information security, as laid down in the Data Agenda Government [14].

20.4.2 Spatial Data Infrastructure in The Netherlands

The Spatial Data Infrastructure (SDI) is a part of the overall public data infrastructure. In The Netherlands, all public parties have established Geonovum as the national geo-standards' development organization (www.geonovum.nl). Geonovum serves as the national SDI executive committee and INSPIRE liaison.

Within the Dutch SDI, much attention is paid to establishing information standards. These have been developed for all themes. The cadastral registers are based on the Information Model Cadastre IMKAD, which is a sector-specific implementation of the "Basic schema for geoinformation – Terms, definitions, relations and general rules for the interchange of information of spatial objects related to the Earth" [15, 16]. All relevant themes have their own NEN3610-compliant information model, for example, Topography (IMTOP), Cables and Pipelines (IMKL), Cultural Heritage (IMKICH), Water (IMWA), and Spatial planning (IMRO), etc. (Fig. 20.24).

This common general information model makes the combined use of data easier and enables data exchange. As the cadastral parcel is one of the layers defined in the EU directive INSPIRE, the cadastral registers are compatible with INSPIRE standards and LADM [1].

The Dutch government has put much effort making spatial data available and accessible. For this, it has established the PDOK platform (www.pdok.nl), which is managed by Kadaster. PDOK distributes (nationwide) datasets (including the cadastral map) through downloads, web services, and

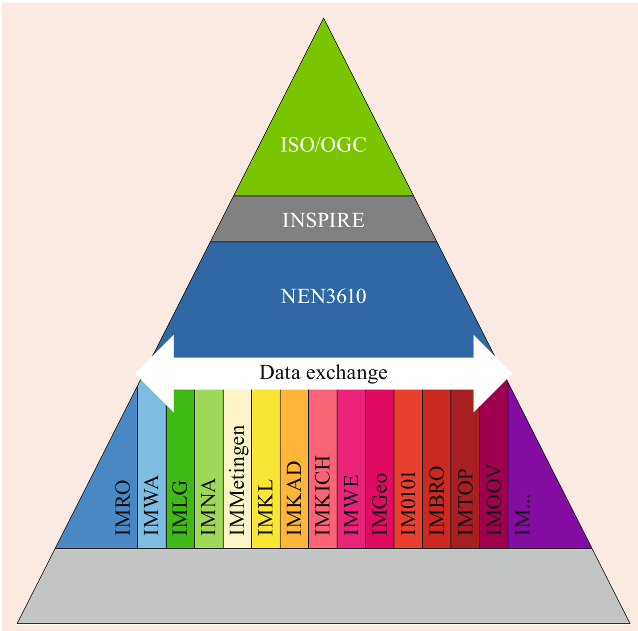


Fig. 20.24 NEN3610: Basic schema for geoinformation – Terms, definitions, relations and general rules for the interchange of information of spatial objects related to the Earth [15]

increasingly via APIs. PDOK also serves as the national INSPIRE portal.

PDOK hosts the National Georegister (NGR; www.nationaalgeoregister.nl), which provides access to and information on spatial information (Fig. 20.25).

20.5 Cadastral and Land Register Data

Cadastral and land register data are available at different levels in different forms.

20.5.1 Public Registers of Deeds

Public registers are registers in which notarial deeds are recorded in the order in which they have been submitted. Public registers are comparable with the land registers kept by courts in other countries. The reason for filing in this order is the importance of the ranking of real rights. The Civil Code (Roman–French law family) assigns two important characteristics to real rights, namely a real right moves with the object, and older real rights have priority over younger real rights. With respect to the latter, the moment of recording can, therefore, be of crucial importance, for instance, regarding legal foreclosure and enforcement.

The public registers were originally kept in analogue format: books with paper deeds, copied to microfiche. Nowadays, all new deeds are entered digitally. As a result of a large scanning project, all historic analogue deeds are digitally available as well.

20.5.2 Cadastre

To enhance public access to the public registers, we extract the essential elements from the deed. These data form the input for the cadastral registers and maps, providing for registers on name, parcel (both administration and cartography), street address, and zip code. In essence, the cadastral registers and maps are auxiliary registers, containing the access keys to the public registers. Both cadastral registers and cadastral maps are 100% in digital format.

In addition to the basic relationship man–right–parcel, there are many attributes. Examples are land use, purchase prices, various legal essential data, and parcel surface area.

20.5.3 Cadastral Map

The cadastral map is maintained in the national grid and shows cadastral boundaries, parcel identifiers, street addresses, contours of buildings, and house numbers. The cadastral boundaries and the parcel identifiers are authentic data on the cadastral maps. The other data are taken from other key registers, mainly the key register on large-scale topography.

The cadastral map provides an overview of the parcels and is, in essence, a cadastral index map, which provides entry into the cadastral registration and access to the information on the cadastral boundaries. It does not provide precise information on the boundary itself.

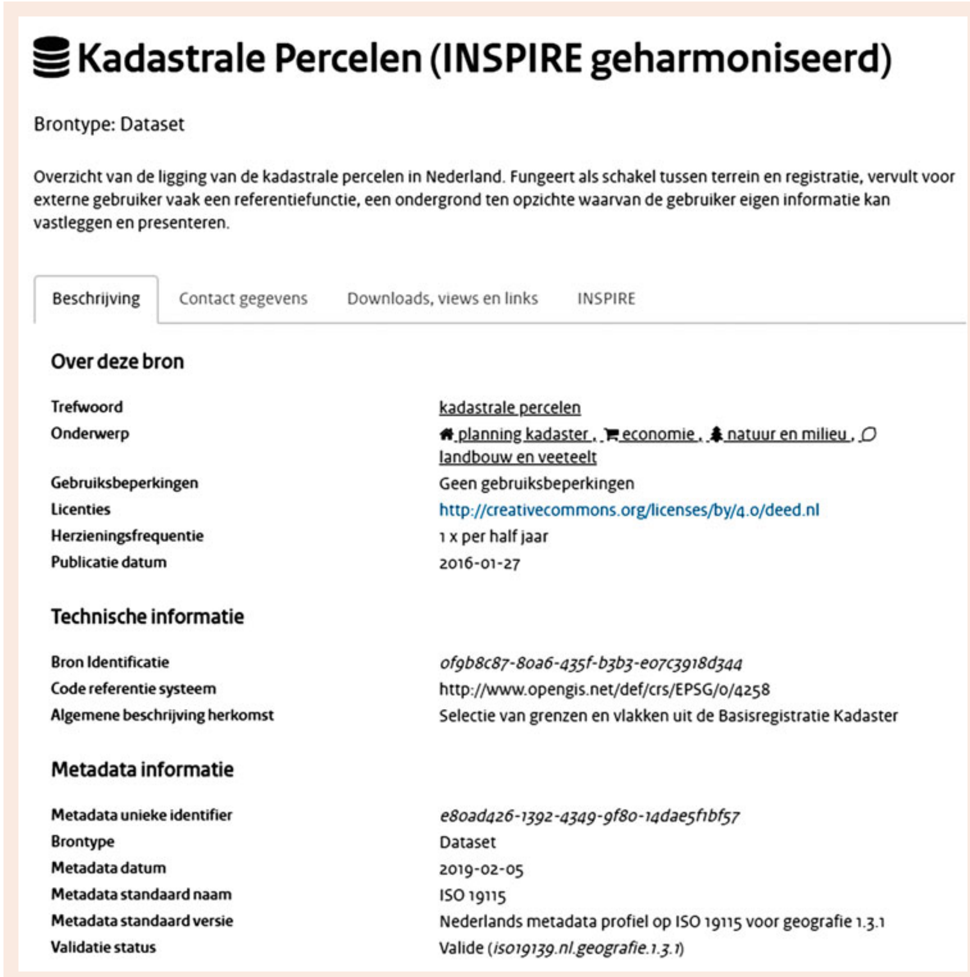
In The Netherlands, the transfer of a (newly defined) property is possible without having surveyed the new parcel beforehand. In order to facilitate this procedure the new parcel(s) is (are) created by drawing up a provisional boundary with the notary. This provisional boundary is depicted in the cadastral map and part of the registration process (Fig. 20.26).

At a later moment in time, the boundary is measured by a cadastral surveyor, and the information is laid down in a (digitally available) protocol, which includes a field sketch with the measurement information.

This sketch and the measurements therein are then used to update the cadastral map (in which the provisional boundary is already shown) with the boundary, as pointed out in the terrain.

Additionally, all existing field sketches have been digitized, so that we have a complete basis for the reconstruction of cadastral boundaries in the terrain.

Fig. 20.25 Sample from the National Georegister on the theme cadastral parcel in INSPIRE



Kadastrale Percelen (INSPIRE geharmoniseerd)

Brontype: Dataset

Overzicht van de ligging van de kadastrale percelen in Nederland. Fungeert als schakel tussen terrein en registratie, vervult voor externe gebruiker vaak een referentiefunctie, een ondergrond ten opzichte waarvan de gebruiker eigen informatie kan vastleggen en presenteren.

Beschrijving | Contact gegevens | Downloads, views en links | INSPIRE

Over deze bron

Trefwoord	kadastrale percelen
Onderwerp	planning kadaster, economie, natuur en milieu, landbouw en veeteelt
Gebruiksbeperkingen	Geen gebruiksbeperkingen
Licenties	http://creativecommons.org/licenses/by/4.0/deed.nl
Herzieningsfrequentie	1 x per half jaar
Publicatie datum	2016-01-27

Technische informatie

Bron Identificatie	*0f9b8c87-80a6-435f-b3b3-e07c3918d344*
Code referentie systeem	http://www.opengis.net/def/crs/EPSG/0/4258
Algemene beschrijving herkomst	Selectie van grenzen en vlakken uit de Basisregistratie Kadaster

Metadata informatie

Metadata unieke identifier	*e80ad426-1392-4349-9f80-14dae5f1bf57*
Brontype	Dataset
Metadata datum	2019-02-05
Metadata standaard naam	ISO 19115
Metadata standaard versie	Nederlands metadata profiel op ISO 19115 voor geografie 1.3.1
Validatie status	Valide (*iso19139.nl.geografie.1.3.1*)

Fig. 20.26 Cadastral map of The Netherlands; cadastral boundaries and parcel numbers (*black*); provisional boundaries (*brown*); contours of buildings (*ground level*) in *red*; house numbers in *red*

Recently, Kadaster embarked on the improvement of the spatial cadastral information (cadastral map and boundary information). We have investigated whether and how information on boundaries can be made available more easily to both professionals and the general public. The chosen option is that the cadastral map is improved in such a way that it provides direct access to boundary information as well.

20.6 Developments in the (Cadastral) Spatial Domain

Cadastral information is becoming increasingly more spatially oriented. From a legal perspective, the delimitation of rights was considered useful to make the outline of rights tangible and visible. In the land market, it was generally the (spatial) object that was pivotal and of which the rights were essential to determine (the level of) control, value, and use. Increasingly, the concept of legal spaces being embraced. In the last decade, the concept of 3-D cadastres witnessed a rapid development, but at the same time, the number of operational 3-D cadastres was limited.

Traditionally, the delimitation of rights has to a large extent been descriptive. Examples in private law are encumbrances and easements. At Kadaster we have investigated how we can easily detect these rights in the deeds and how we can make the delimitation of the rights visible. A good example is the so-called “right of way”, which does not usually concern the whole parcel but, in practice, gives right of way to a certain part of the property.

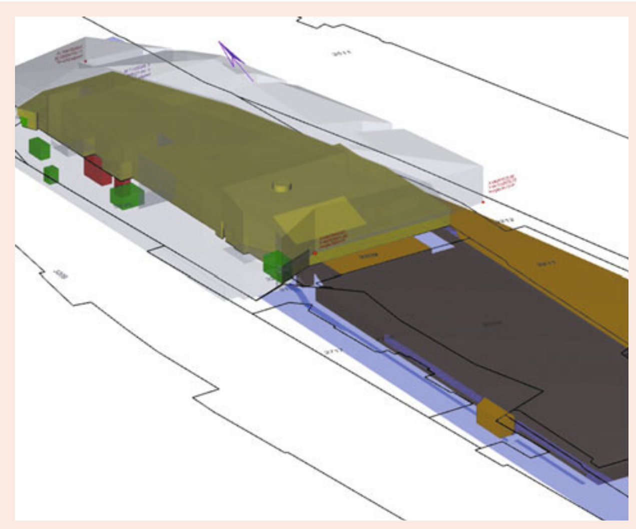


Fig. 20.27 Delft railway station in 3-D; *gray*: freehold city hall; *purple*: right of superficies railway tunnel; *brown*: right of superficies user terminal; *light green*: right of superficies station hall; *red*: right of superficies stairs and elevator; *green*: right of superficies technical installations

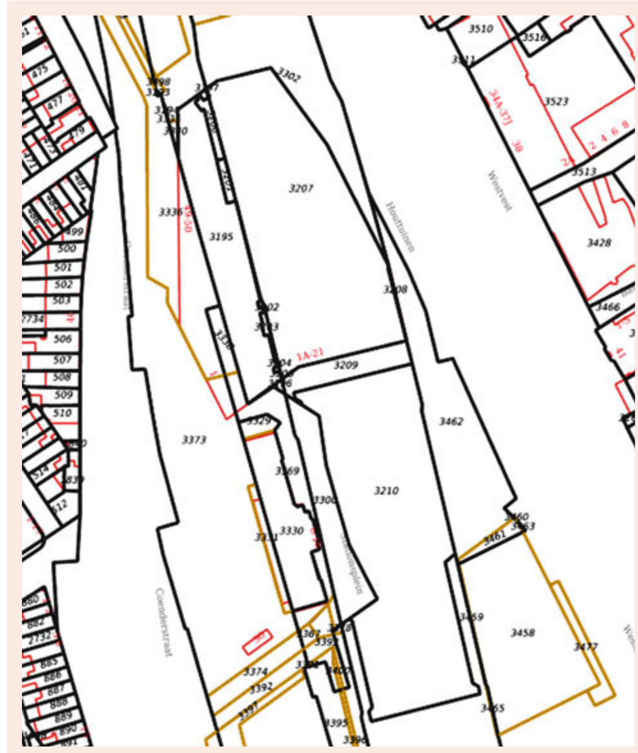


Fig. 20.28 Situation as depicted in Fig. 20.26 represented on the conventional cadastral map

In public law, we enable the description of public restrictions by contours, whereas it was traditionally usual to list all the parcel to which the restriction related. By making registration by spatial means possible the information on these public and private rights is resilient to changes in the property object and the parcellation.

Furthermore, we are witnessing an ongoing integration of data sources. A good example is the system of key registers presented earlier. This does not only require technical and semantic interoperability; changes in the real world (such as physical objects) can have legal and/or administrative consequences or vice versa. To obtain an information system where these three occurrences are in sync also requires organizational and sometimes even legal interoperability. In practice, these changes are not concurrently visible in the various datasets at the moment.

20.6.1 Case: 3-D Cadastre

Increasing complexity in the built environment requires that also the transparency of rights, restrictions, and responsibilities is improved. In The Netherlands, with the help of architects, notaries, and academia, we have experimented with visualizing and registering rights in space for the new underground railway station in Delft with the new city hall above ground level. The end result was the creation of a 3-D pdf, which was attached as an appendix to the deed and was registered in the land registry (Fig. 20.27; [17]).

This solution provides the insight that is required in cases where we have to deal with legal spaces – an insight that the conventional cadastral map cannot provide (Fig. 20.28).

Following this development we are now investigating how we can model legal spaces concerning apartment rights. Apartment rights are currently delimitated by 2-D floor plans. A first step is to devise a methodology to automatically extract (3-D) legal spaces from the building model itself. The basis is the IFC (BIM) (Industry Foundation Classes, Building Information Modeling) data model, which does not support legal spaces. By enriching IFC BIM with legal space by a cadastral information user-defined property set, we believe that we can find a solution that fits the existing legal framework and current practices (Fig. 20.29) [18].

20.6.2 Case: Delimitation of Public Restrictions

In The Netherlands the registration of public restrictions is being renewed. All restrictions will be registered in the key register cadastre. As a main change to the current situation where restrictions are registered based on parcels, we will introduce the registration of restrictions by contours

Fig. 20.29 First results of BIM Legal: extracting legal spaces using the IFC data model

Fig. 20.30 Delimitation of public restrictions by means of a contour as a separate legal theme, where only a part of a parcel is affected

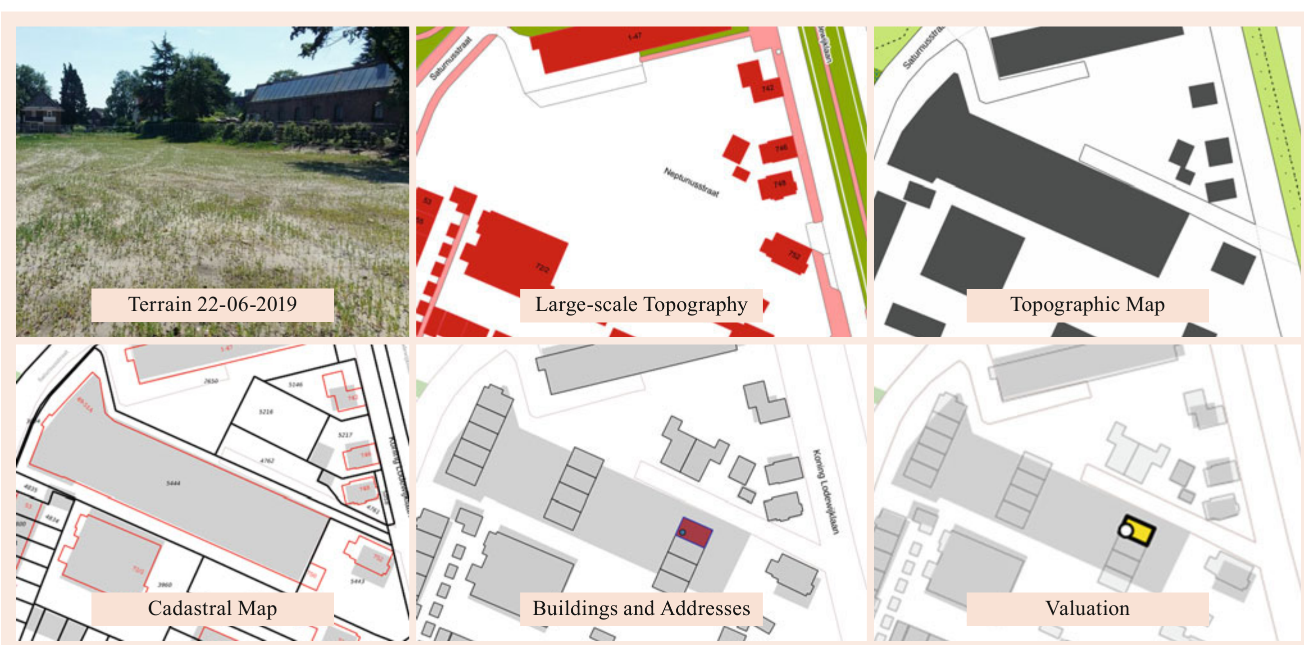


Fig. 20.31 Example of the status of a terrain in different key registers at a reference date

(Fig. 20.30), whereby the contour determines the spatial extent of the restriction, irrespective of changes in the parcellation at a later moment in time. This provides a much better insight in the legal situation to parties in the land market.

20.6.3 Case: Working Towards an Integrated Information Model

The demand for integrated information on properties and spatial objects is growing. People expect that information on an object (e.g., a property or a building) is up to date, correct, and complete. We live in a world where we have been organized around processes (be it property transfer, building permits, or valuation). The result is that data relate to each other in the long run (explicitly by updates or implicitly via location) but are not integrated at a certain moment in time and thereby providing a "common operational picture". An example of such a situation is depicted in Fig. 20.31, where we visualize the content of separate information sources. It concerns a terrain that has been cleared and where a housing development has received a building permit. Each registration by itself is up to date according to its specifications, but at the same time the differences between the registration are obvious to the end user.

Linked data is an approach to publishing data that puts linking at the core of the data representation. In this manner, changes in an object propagate automatically to all the different registrations. Linked data requires that the links be (explicitly) defined and permanently maintained. Here, we can create a so-called knowledge graph where interlinked data are used for relationship definition and metadata. We are now looking into methodologies where we can also integrate implicit relations into the design. In such a way, we hope to evolve to a sustainable, linked, and up-to-date system of registrations.

20.7 Concluding Remarks

The spatial aspect of rights has gained momentum and makes the cadastre and land register more accessible, usable, and versatile. At the same time, (spatial) cadastral information is more and more embedded in the overall (spatial) data infrastructure. This leads to a convergence in modeling aspects and even brings different worlds, such as building and geoinformation, closer together. On the one hand, this makes the cadastral spatial information less special, and, on the other, cadastres and land registers are updated on a daily basis and, thereby, provide a foundation for a sustainable (spatial) data infrastructure.

References

1. ISO 19152:2012: Geoinformation – Land Administration Domain Model (LADM). International Organization for Standardization, Geneva (2012)
2. EC: Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). European Commission, Brussels (2007). http://inspire.jrc.ec.europa.eu/
3. ISO 19135-1:2015: Geographic Information – Procedures for Item Registration – Part 1: Fundamentals. International Organization for Standardization, Geneva (2019)
4. ISO 19111:2019: Geographic information – Spatial Referencing by Coordinates. International Organization for Standardization, Geneva (2019)
5. AdV-WG Information and Communication Technology: Documentation on the Modelling of Geoinformation of Official Surveying and Mapping in Germany (GeoInfoDoc) – Version 7.1, Working Committee of the Surveying Authorities of the States of the Federal Republic of Germany, AdV: http://www.adv-online.de
6. ISO 19103:2015: Geoinformation – Rules for Application Schemas. International Organization for Standardization, Geneva (2015)
7. ISO 19101-1:2014: Geoinformation – Reference Model – Part 1: Fundamentals. International Organization for Standardization, Geneva (2014)
8. ISO 19109:2015: Geoinformation – Rules for Application Schemas. International Organization for Standardization, Geneva (2015)
9. ISO 19107:2019: Geoinformation – Spatial Schema. International Organization for Standardization, Geneva (2019)
10. ISO 19118:2011: Geoinformation – Encoding. International Organization for Standardization, Geneva (2011)
11. Open Geospatial Consortium: Geography Markup Language (GML) Implementation Specification. OGC, Wayland (2019). http://www.opengeospatial.org/specs/
12. Ministry of the Interior and Kingdom Relations Digital Government website, https://www.nldigitalgovernment.nl/dossiers/base-registers-and-system-standards/
13. Logius Catalogue of System of Key Registers (in Dutch), https://www.stelselcatalogus.nl/
14. Ministry of the Interior and Kingdom Relations: NL Digitaal Data Agenda Governement (2019). https://www.nldigitalgovernment.nl/wp-content/uploads/sites/11/2019/04/data-agenda-government.pdf
15. NEN 3610:2011: Basic Schema for Geo-Information – Terms, Definitions, Relations and General Rules for the Interchange of Information of Spatial Objects Related to the Earth (2011). https://www.nen.nl/NEN-Shop-2/Standard/NEN-36102011-nl.htm
16. van den Brink, L., Janssen, P., Quak, W., Stoter, J.: Towards a high level of semantic harmonisation in the geospatial domain. Comput. Environ. Urban Syst. **62**, 233–242 (2017)
17. Stoter, J., Ploeger, H., Roes, R., van der Riet, E., Biljecki, F., Ledoux, H.: First 3D cadastral registration of multi-level ownerships rights in the Netherlands. In: van Oosterom, P., Dimopoulou, E., Fendel, E. (eds.) Proc. 5th Int. FIG Workshop 3D Cadastres, pp. 491–504. International Federation of Surveyors (FIG), Copenhagen (2016)
18. Meulmeester, R.: BIM Legal: Proposal for Defining Legal Spaces for Apartment Rights in the Dutch Cadastre Using the IFC Data Model. MSc thesis. Delft University of Technology, Delft (2019)
19. Object Management Group: Unified Modeling Language – A specification Defining a Graphical Language for Visualizing, Specifying, Constructing, and Documenting the Artifacts of Distributed Object Systems. OMG, Needham (2019)

Markus Seifert Markus Seifert is a surveying and geographic information engineer. His area of research is data modeling, standards for spatial data infrastructures, and digital cadastral systems. He is Chair of the Standardization Working Group Cartography and Geoinformation at the German Standardization Institute (DIN) and Representative for the German Mapping Agencies to national and international standardization organizations for geographic information (SDI Germany, OGC, ISO/TC 211, CEN/TC 287).

Martin Salzmann Martin Salzmann is the strategy lead with the Cadastre, Land Registry and Mapping Agency (Kadaster) of The Netherlands. He holds a PhD from Delft University of Technology. In the past, Martin worked extensively in the fields of quality assurance of cadastral surveying and mapping, information strategies, and marketing before moving into the realm of strategic planning and eGovernment.

Building Information Modeling

21

Jörg Blankenbach and Ralf Becker

Contents

Abstract

Digitalization represents a global megatrend that is leading to fundamental changes in numerous areas of society, technology, and production – including the construction industry (AECO, architecture, engineering, construction and operation). A key element in the digital transformation of the construction industry is building information modeling (BIM). At its core, BIM aims at the holistic, consistent acquisition, management, and exchange of all relevant building information based on digital models. As an integral working method, BIM has a fundamental impact on design, construction, and operational processes and thus addresses all the stakeholders involved in the lifecycle of buildings. Even though digital construction and operation is not yet a standard in practice, BIM is developing rapidly and is increasingly being applied in the construction industry.

This chapter introduces the BIM method and the link to geospatial information modeling (GIM). The basics and related processes are explained, and use cases and examples are presented. Standardization aspects are dealt with, and the state of implementation at national and international levels is outlined. One focus lies on the complementary to GIM and approaches of BIM-GIM interoperability.

Keywords

building information modeling (BIM) · digital design and construction · BIM processes · IFC

Digitalization and automation are global megatrends. They are also among the core future challenges of the construction industry (the AECO industry). A significant contribution in this context is the introduction of the new methodology of building information modeling (BIM). BIM is synonymous with digital planning, construction, and operation, and promises efficiency gains in the acquisition, management, and exchange of building-related data throughout the entire building lifecycle.

The planning, construction, and operation of a building are highly interdisciplinary tasks. Optimal value creation therefore requires seamless digital process chains as well as seamless exchange of information between all parties in-

J. Blankenbach (✉)
Geodetic Institute and Chair for Computing in Civil Engineering and Geo Information Systems, RWTH University Aachen
Aachen, Germany
e-mail: blankenbach@gia.rwth-aachen.de

R. Becker
Geodetic Institute and Chair for Computing in Civil Engineering and Geo Information System, RWTH Aachen University
Aachen, Germany
e-mail: ralf.becker@gia.rwth-aachen.de

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_21

volved. Thus, the use of integrated digital building models, open standards, and data management in shared data environments throughout the entire building lifecycle are key aspects of BIM that enable optimal collaboration between different disciplines.

Many tasks in the building lifecycle require the involvement of spatial information (geospatial data). For example, real estate data, city and landscape models, or terrain models are needed for the planning of new buildings. Due to the increasing use of BIM, digital data from the geospatial information modeling (GIM) domain has to be integrated into BIM. In the future, BIM and GIM will no longer stand side-by-side in the building lifecycle; for an optimal planning, construction, and operating process, they should be seen as complementary.

This chapter introduces the BIM method, identifies the differences between BIM and GIM, and highlights the possibilities for establishing interoperability between the BIM and GIM data models.

21.1 Introduction to Building Information Modeling

The basic idea of BIM – the holistic digital modeling of buildings – is not new and has long been discussed in science and research, e.g., in [1]. The US National Institute of Building Sciences (NBIM) described BIM in 2007 as a product, collaborative process, and facility lifecycle requirement [2]. Currently, NBIM defines BIM as "a digital representation of physical and functional characteristics of a facility. As such, it serves as a shared knowledge resource for information about a facility, forming a reliable basis for decisions during its lifecycle from inception onward" [3].

The core element of BIM is a shared database in the form of digital, holistic, and uniform building models (Fig. 21.1a). In defined and standardized processes, those involved in the process feed this digital database and access it equally. This process is not limited to the planning phase, but ideally extends over the entire lifecycle, including construction, operation, and management (e.g., facility or maintenance management) as well as renovation or demolition (Fig. 21.1b; [4]).

In addition to the three-dimensional geometric or building structure concerning information, attribute data for describing technical or functional aspects, time and cost management, etc., are also integrated into BIM, so that a multidimensional database is created. The use cases of BIM in the building lifecycle are correspondingly diverse, and include:

- Planning:
  - Visualizations, as well as the derivation of sections and plans
  - Component list/bill of material; area, volume, and quantity determinations; technical specification
  - Model-based cost estimation and time management
  - Coordination, clash detection
  - Simulations/variant studies
- Construction:
  - Creation of execution plans; site management and logistics
  - Construction (progress) documentation/control
  - Defect management and documentation
- Handover/commissioning/approval:
  - As-built documentation
  - Billing of construction works

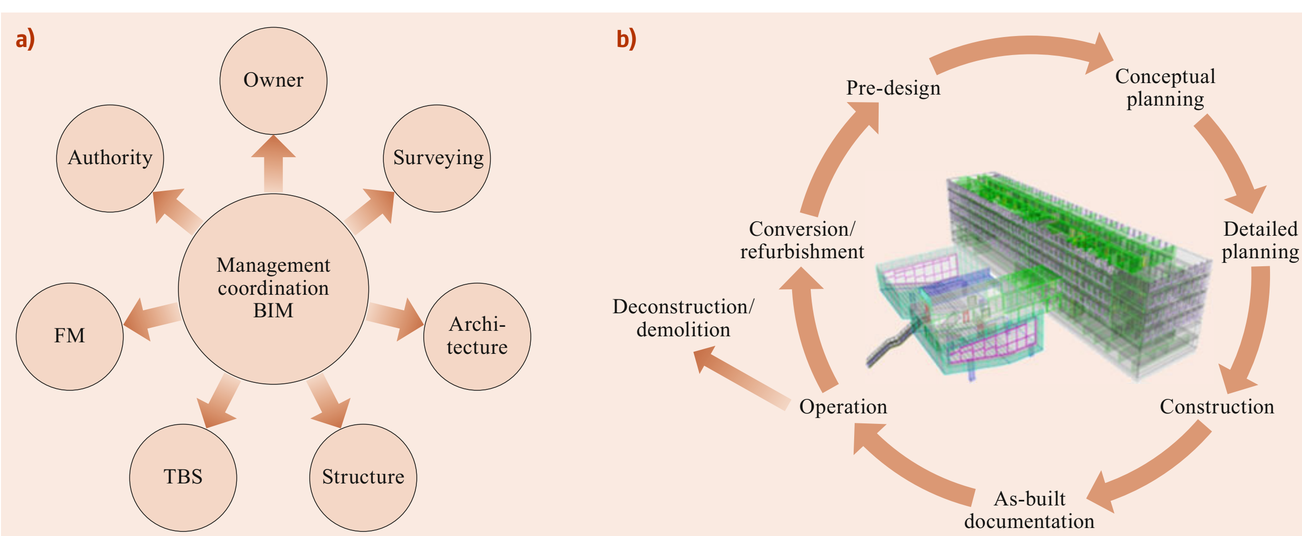


Fig. 21.1 **a** BIM and the trades involved in construction (FM = Facility Management, TBS = Technical Building Services). **b** BIM in the building lifecycle

Fig. 21.2 New Trier lock: colored point cloud of a laser scanner point cloud (**a**), digital as-built model (**b**)

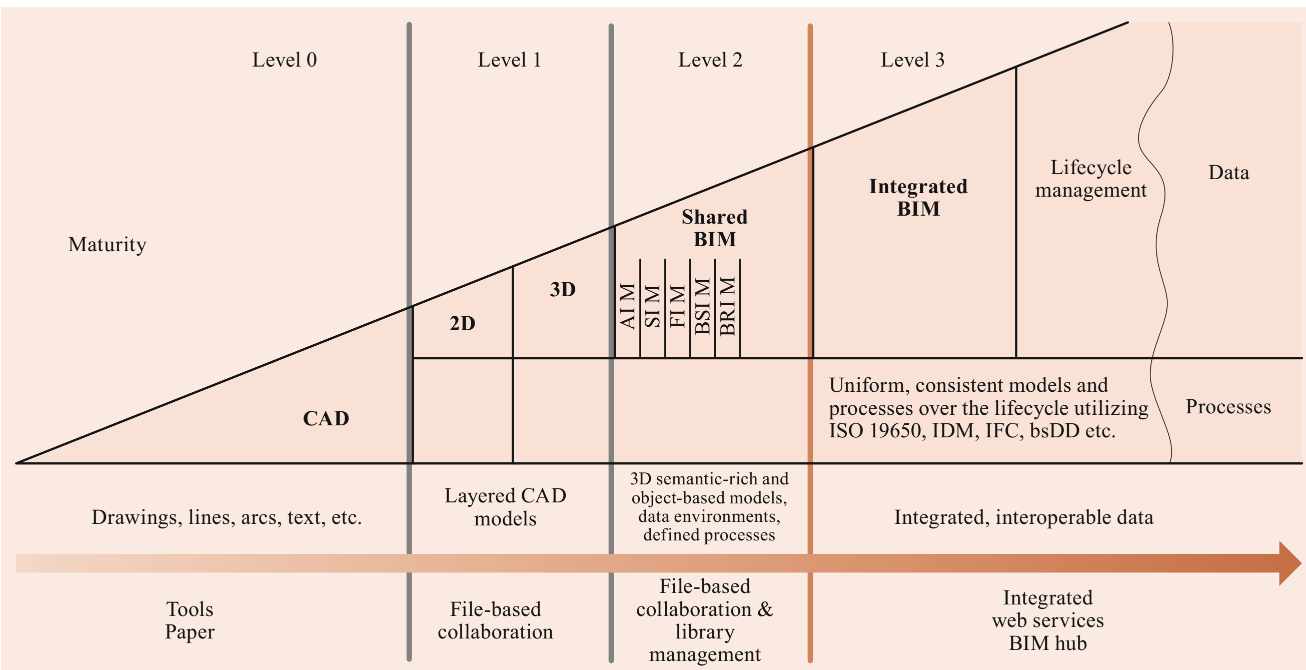


Fig. 21.3 BIM maturity levels (AIM = Asset Information Model, SIM = System Information Modelling, FIM = Facility Information Modeling, BSIM = Building Service Information Model, BRIM = Bridge Information Model, IDM = Information Delivery Manual, IFC = Industry Foundation Classes, bsDD = buildingSMART Data Dictionary) (adapted from: [5])

- Operating phase:
  – As-is documentation
  – (Computer-aided) facility management (CAFM)
  – Maintenance management
  – Deconstruction/dismantling planning/recycling.

BIM originates from building construction but is increasingly also being utilized in infrastructure construction. Occasionally, therefore, the term CIM (construction information modeling) is used in analogy to BIM.

An example of the use of BIM in the lifecycle is shown in Fig. 21.2. An as-built model of a newly constructed structure (a ship lock) was derived from laser scanner and image data in order to enable a comparison with the planning and to be used in the operating phase as a *digital twin* of the structure that was actually built (e.g., for maintenance management).

Generally, BIM assumes that working together on a holistic, consistent database improves communication and coordination between disciplines. Thus, discrepancies and errors should be detected and corrected in the early planning phases, and information gaps or losses should be avoided.

However, the introduction of BIM represents a paradigm shift in the AECO industry that will not succeed immediately. Therefore, the introduction of BIM involves the incremental digitization of data, data storage, data exchange, and (collaborative) processes. The working method changes from the use of analog drawings to file-based data exchange and, finally, to data clouds. Likewise, the classical, often two-dimensional, CAD-based approach is being replaced by semantic-rich, object-oriented work in 3-D. The transition to digital exchange and digital processes requires the introduction of standards for data exchange that are as open as possible, the use of component libraries, and the standardization of processes. The introductory history of BIM is often subdivided into so-called maturity levels, e.g., by the British BIM Industry Working Group (Fig. 21.3; [5]).

Also related to the adoption of BIM is the degree and nature of collaborative working between stakeholders, reflected in the distinction between closed/open and little/big BIM, respectively, and their permutations. The term *closed* refers to the use of only one software, in contrast to *open*, which involves software from various manufacturers and open formats for data exchange. *Little* refers to the application of BIM for individual trades only, as opposed to the utilization of BIM by several or, ideally, all specialist disciplines (*big*).

21.2 Modeling

Traditionally, the design and planning of buildings is conceived and documented in two-dimensional drawings and sections. Today, they are usually created in CAD systems and consist of geometric primitives such as points, lines, and surfaces, which receive their meaning solely through layer selection and graphical characteristics. Further information is, at best, added by introducing labels that have no direct relation to the object except placement. Depending on the CAD software used, floor plans and sections can be combined into three-dimensional visualizations.

21.2.1 Component-Oriented Modeling

BIM's modeling paradigm, unlike the drawing-based CAD approach, is based on an object-oriented method and semantic data models. The basic information carrier is the building object or component, e.g., a wall, beam, or column, which has various properties: semantics, attribute data, three-dimensional geometry, and relations to other building elements (neighborhood, group membership) (Fig. 21.4).

The attribute data may include physical, functional, technical, or descriptive information. Other data such as time and cost plans or documentation can also be linked to the components in this way. Figure 21.5 shows an example of a building section taken from an architectural model.

Thus, based on the component-oriented models, a variety of analyses and simulations, such as the creation of material lists, cost planning, or mass/volume calculations (see the BIM use cases in Sect. 21.1), become possible.

21.2.2 Level of Development and Level of Accuracy

The design and planning of structures are done step-by-step, starting with predesign and conceptual planning and ending with execution planning. This applies equally to GIM- and BIM-based planning in civil engineering and building construction. This begins with conceptual studies or route variants, which, for example, initially only create a visual impression of how the designs fit into the landscape or cityscape. For this, the digital models and thus the components initially require only coarse detailing. With refined planning (submission of the documents for the planning application or planning approval, the determination of masses or quantities, cost estimation, etc.), the demands on the detailing for both the geometric and the attributive information increase. To meet the different requirements for the detailing of model objects in planning, the concept of the *level of development* (LOD) has been introduced (this is referred to as LOD_{BIM} in the following). Examples can be found in different specifications, e.g., in [6] for Australia (Fig. 21.6)

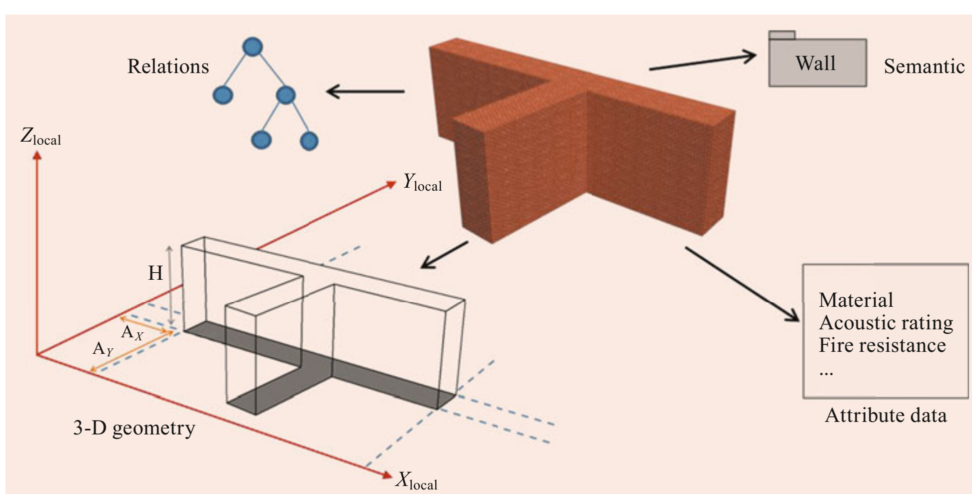


Fig. 21.4 Semantic data models – objects with attributes and relations

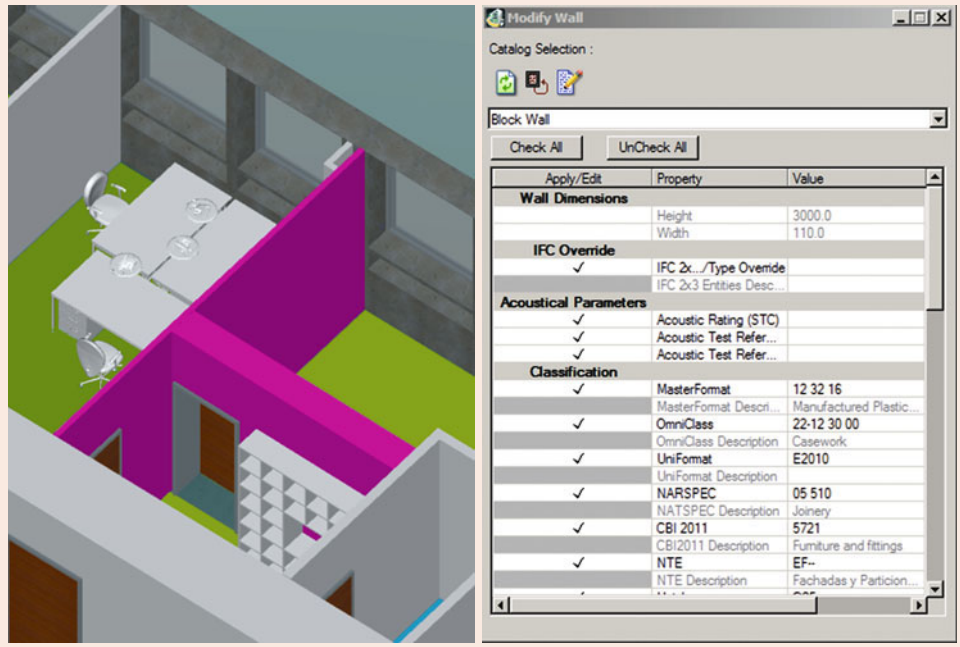


Fig. 21.5 Component "wall" with attribute data in BIM software (Bentley AECOsim BuildingDesigner)

or the US LOD specification [7]. The final $\mathrm{LOD}_{\mathrm{BIM}}$ is often called *as-built*. In contrast to the levels of the planning phases, as-built reflects the state actually realized, which, in practice, often differs from the execution planning, and can only be achieved by comparison with the final planning status on site. Differentiation into geometric and attributive parts can be found in the terms *level of geometry* (LOG) and *level of information* (LOI). This $\mathrm{LOD}_{\mathrm{BIM}}$ should not be confused with the term *level of detail* (also abbreviated to LOD, but referred to as $\mathrm{LOD}_{\mathrm{GIM}}$ in the following) used with city models for geospatial information modeling (Sect. 21.4.3).

The $\mathrm{LOD}_{\mathrm{BIM}}$ must be distinguished from the *level of accuracy* (LOA). LOA defines the geometric accuracy with which a component is to be surveyed onsite and – to be handled separately – the geometric accuracy of the model (as-built or as-is model) subsequently derived from the surveying data. The data capturing methods and instruments as well as the modeling software and technology used should be selected based on this. The US Institute of Building Documentation (USIBD) has published the *USIBD Level of Accuracy Specification Guide* [8], which is divided into five classes.

21.2.3 Aspect or Domain Models

The BIM method is based on the creation of a digital building model. However, when several disciplines are involved, each specialist planner usually creates his own domain-specific model (domain or aspect model). Each aspect model takes the specific requirements of specialist planning into account, which are also expressed in different software tools. Only the merged aspect models result in the overall model (coordination model, Fig. 21.7).

As an example of a building construction project, the following aspect models are possible according to [10]:

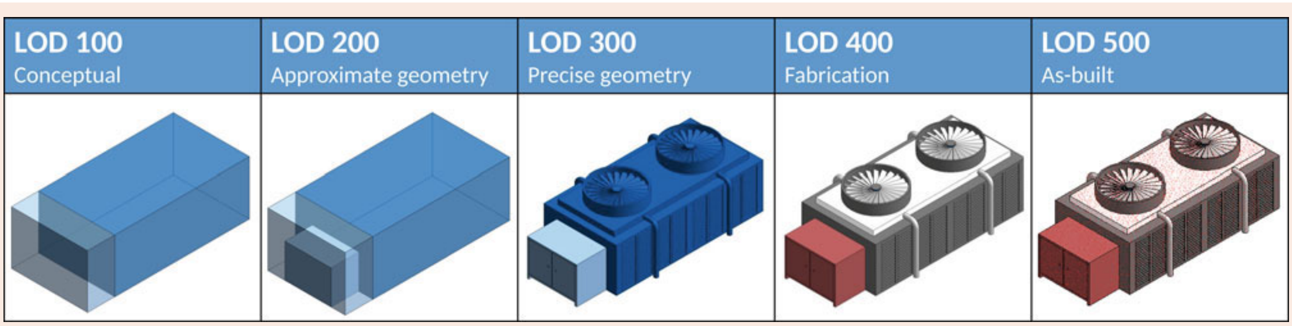


Fig. 21.6 $\mathrm{LOD}_{\mathrm{BIM}}$ (adapted from [6])

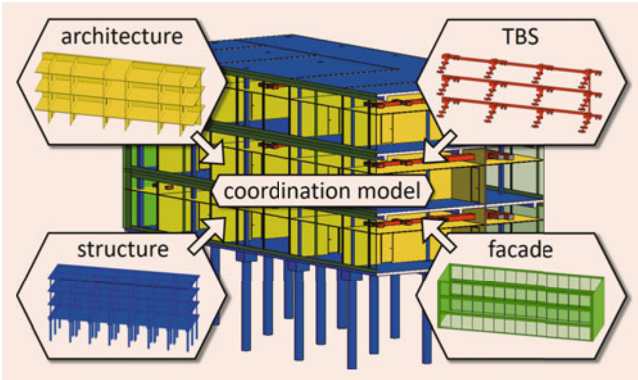


Fig. 21.7 BIM – coordination and aspect models

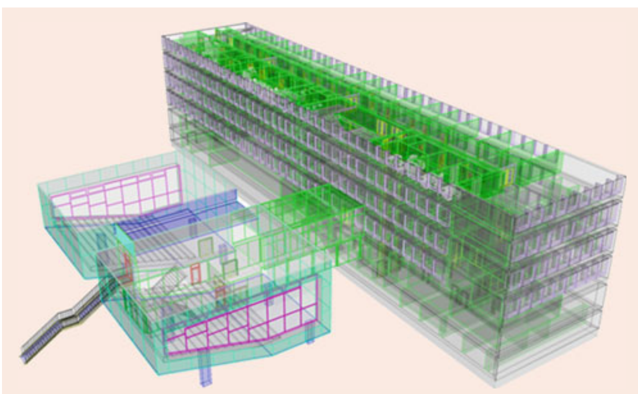


Fig. 21.8 BIM architecture model (visualization) of the faculty building for the civil engineering of RWTH Aachen University

- Environment model (terrain model, vicinity from a city model)
- Mass model (urban classification)
- Architectural model (at various LOD_{BIM}) (Fig. 21.8)
- Structural model (at different LOD_{BIM})
- Technical building services models (TBS) (at different LOD_{BIM} and in different disciplines)
- Site equipment model
- Construction process model (4-D model)
- Construction and assembly model
- Construction handover or documentation model
- CAFM model.

With the aggregation of the aspect models into a coordination model, geometric and technical evaluation of domain-specific planning becomes possible – in order to detect clashes or rule violations, for example. Model checking information can be exchanged via the standardized BIM Collaboration Format (BCF) [11], which ensures software-independent data exchange.

21.2.4 Industry Foundation Classes (IFC)

The exchange and storage of building model data can be accomplished with different file formats. In addition to the mostly proprietary data formats of software manufacturers, there are Industry Foundation Classes (IFC) – an open, manufacturer- and software-independent data model for BIM that can also be used for data exchange [13]. In IFC, both the

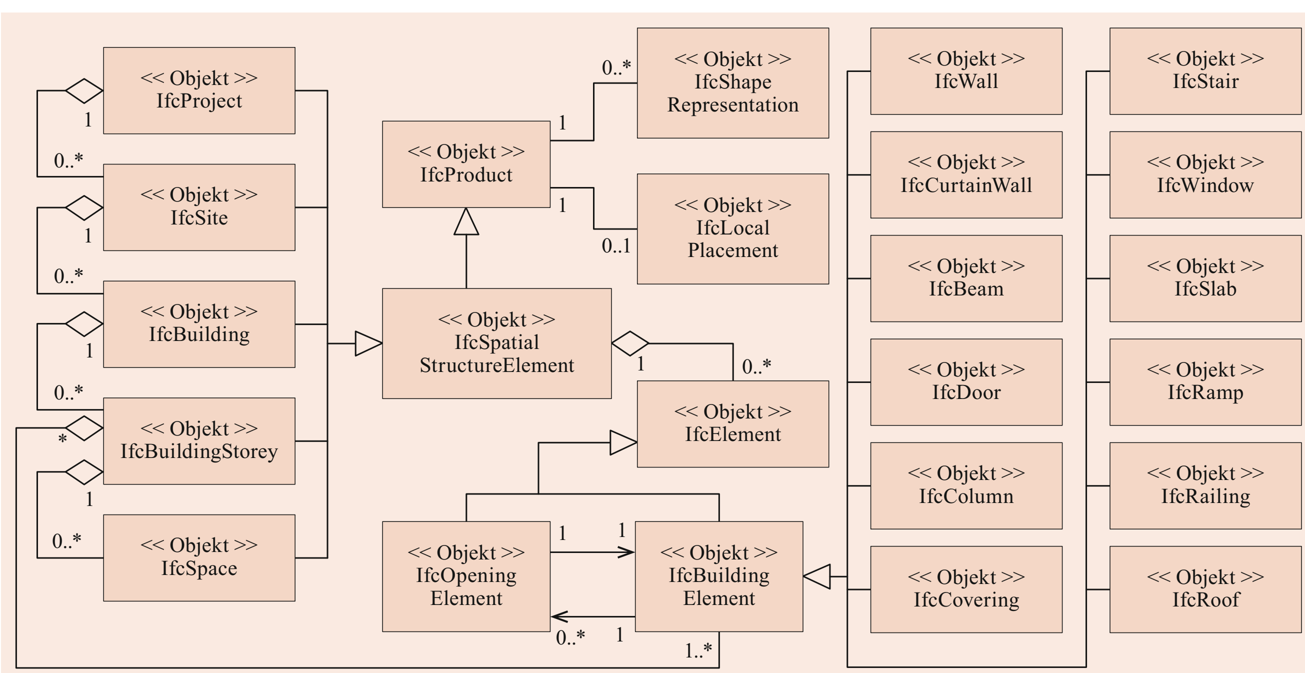


Fig. 21.9 Coarse sketch from the IFC building model (source: [9])

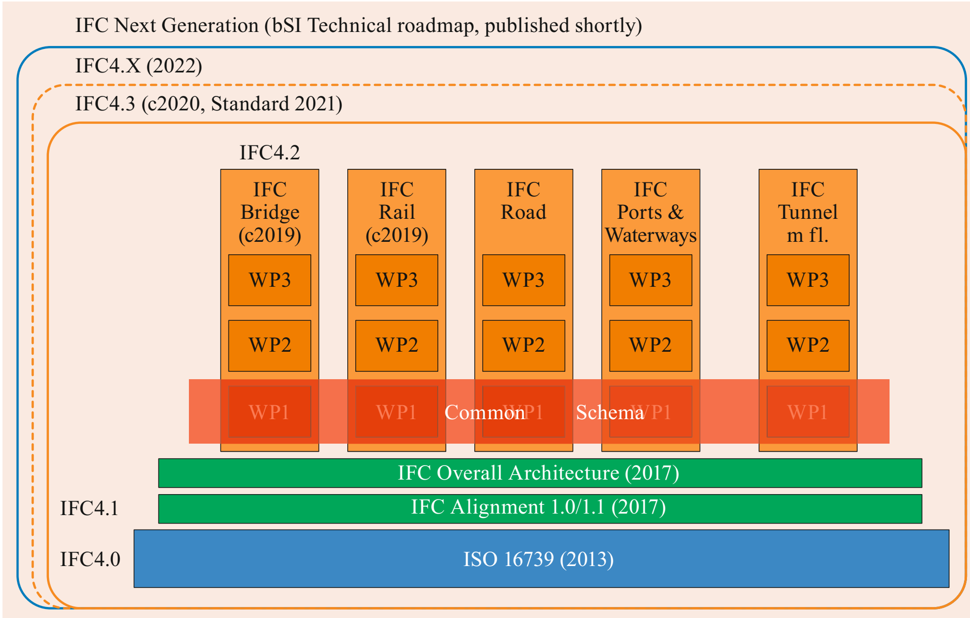


Fig. 21.10 IFC infrastructure extensions (source: [13])

project structure and the model elements can be described along with their geometric and attributive properties as well as the relations between the components (relations). IFC is currently being developed by the nonprofit organization buildingSmart International (it was previously developed by buildingSmart's predecessor organization, AIA). Figure 21.9 shows a coarse extract from the IFC building model.

The IFC information model is hierarchical and based on the data modeling language EXPRESS. Storage and exchange are usually done using the ASCII-based STEP file format, whereby an output in XML (IFCXML) or JSON (IFCJSON) is also possible [14].

Initially designed for building construction, IFC is now standardized as ISO 16379 in the IFC2x3 and IFC4 versions. Due to the increasing use of BIM in infrastructure construction, intensive work is currently being carried out to extend the standard to all areas of infrastructure (rail, roads, bridges, tunnels, ports & waterways). For this purpose, common aspects of all types of infrastructure (e.g., alignment) were introduced with IFC version 4.1. Based on this, cross-discipline aspects (IFC4.2) and specific component classes are defined, respectively are under development, for the respective construction types. Currently (November 2021) they are published as candidate standard (IFC4.3 RC2) for bridges, rail, road and ports & waterways (Fig. 21.10).

21.3 Collaboration Processes

Central components of working with BIM are cooperation and transparent data exchange between the participants. The associated processes require both an organizational (management process) framework and an IT (information process) framework for coordination and cooperation (Fig. 21.11).

These are internationally standardized in ISO 19650, which describes concepts and principles for the successful management of information with a degree of maturity that can be described as "BIM according to ISO 19650." This is applicable to the entire lifecycle of a building, including predesign, planning, construction, operation, maintenance, refurbishment, repair, and demolition/deconstruction.

A data exchange procedure that is uniform and as simple as possible is useful for collaboration. The structures used for information classification and object-oriented information exchange should therefore be standardized, e.g., according to the rules of ISO 12006-2 (*Building Construction—Part 2: Framework for Classification*) or ISO 12006-3 (*Part 3: Framework for Object-Oriented Information*).

21.3.1 Organizational Collaboration

At the organizational level, the BIM process begins with the so-called employer's information requirements (EIR), in which the client has to inform the potential contractors of their requirements (e.g., the LOD_{BIM}, milestones, data formats, the coordinate systems and project reference point to be used, and the preliminary surveying needed are defined in the client's call to tender his requirements). In the course of awarding the contract, the BIM execution plan (BEP) between the client and the contractor is defined. This describes the process of producing the required data and defines all the roles, functions, processes, interfaces, interactions, and technologies to be used. At the end of the award process, the master information delivery plan (MIDP) sum-

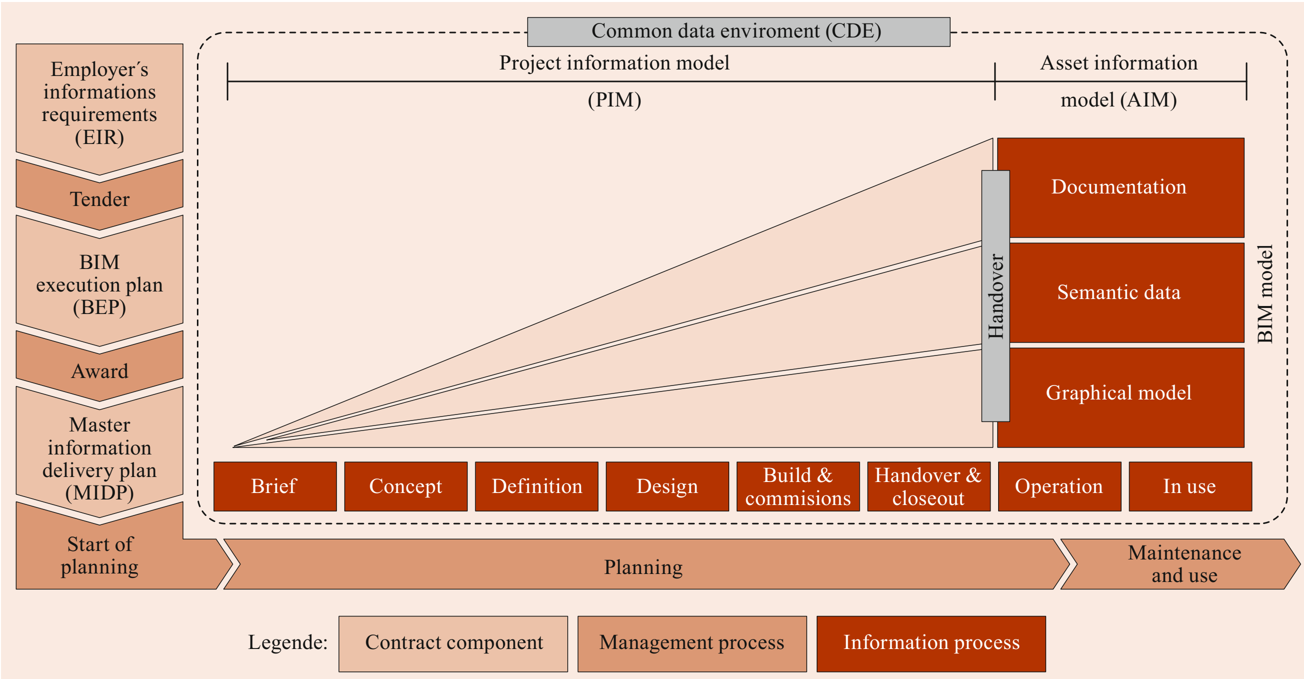


Fig. 21.11 BIM – the process (source: [15])

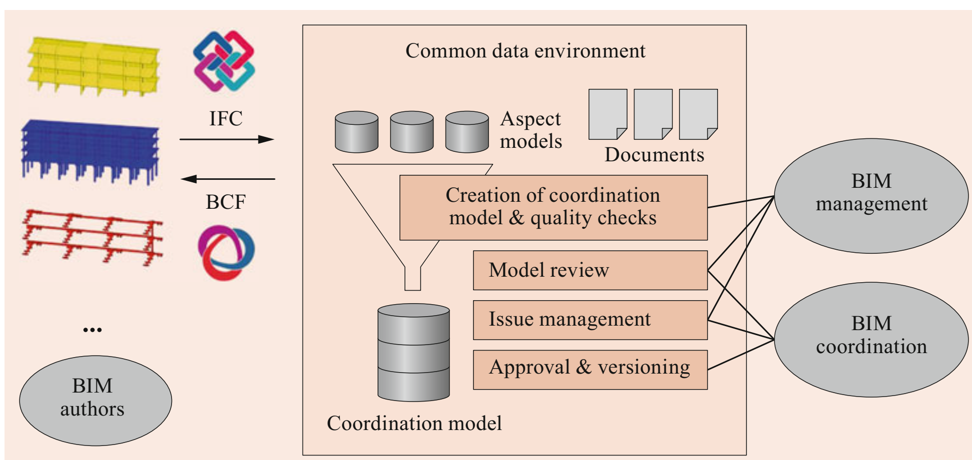


Fig. 21.12 Common data environment (CDE): approval, versioning, and model reviews (BCF = BIM collaboration format)

marizes the information to be provided. buildingSmart has developed a framework, the information delivery manual (IDM), for uniformly describing processes and data transfer points.

21.3.2 Collaboration Platform

Technologically, collaboration is based on a central, common project platform, a data space (common data environment, CDE) in which all participants transfer their aspect models and data for model-based coordination. By aggregating the different aspect models in a coordination model in the CDE, it becomes technically feasible to perform independent geometric and professional quality assurance checks (e.g., clash detection) and provide feedback to the specialist planners (e.g., using BCF; Sect. 21.2.3). Approval and versioning of the models as well as model reviews can also be done via the CDE (Fig. 21.12).

The information stored in the CDE should be understood and linked to by all participants.

This requires coordination regarding:

- The data formats of the stored information
- The exchange formats
- The structure of the information model
- The structure and classification of the data
- The attribute names for metadata.

Ideally, the CDE should not be limited to the planning and construction phase for buildings using the project information model (PIM) but should also extend to the asset information model (AIM) in the operational phase, in line with BIM's objectives. Here, the handover between PIM and AIM is of crucial importance, since value-added reuse of the model in the operational phase can only succeed by comparing and eventually adapting the planning model (as-planned) to the entity that is actually built (as-built).

21.4 BIM vs. GIM

In many cases, due to the complementarity of BIM and GIM, geospatial data (e.g., city and landscape models, terrain models, and environmental data) are also needed in the building lifecycle (e.g., for planning purposes). This increasingly requires the linking and exchange of data between GIM and BIM. For example, BIM planning models can be combined with city models for variant visualization, for shadow and visibility analyses, or to simulate emissions (noise, dust) in geographic information systems (GIS). The results of these analyses can then be used in the BIM world to refine the planning.

Although BIM and GIM have synergies, not least since both domains have overlapping content types and typically rely on semantic data models, their perspectives are fundamentally different. While GIM is characterized by a more generalized view for modeling the *existing reality*, BIM offers a detailed view of the *planned reality* (Fig. 21.13).

These different perspectives lead to some significant differences in modeling paradigms [17].

21.4.1 Modeling Paradigm and Geometry Presentation

The BIM method originated as a building design instrument. The created virtual model is only transferred to reality later in the course of construction. According to the approach used in planning, the geometry is described implicitly as a result of the design process. Therefore, volumetric primitives such as cuboids and cylinders are often combined to represent individual components, and are described parametrically (e.g., by length, width, height). Complex models can be created by the constructive solid geometry (CSG) technique, which involves applying Boolean operators such as union, difference, and intersection to the primitives or using the extrusion and sweep techniques (Fig. 21.14a).

GIM, on the other hand, has its origins in the surveying and mapping of the real world. Surveying typically detects the visible, i.e., the surfaces of the elements, from which a digital

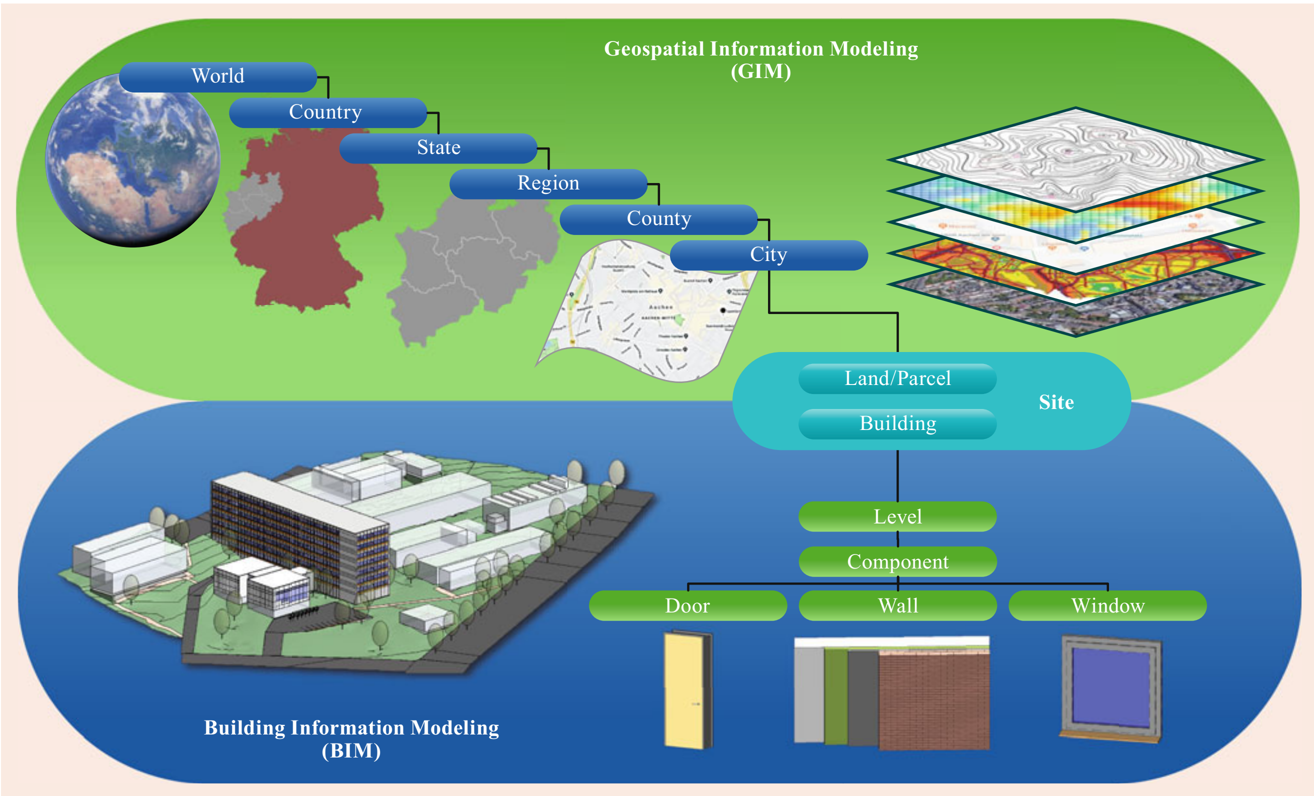


Fig. 21.13 Information relationships between BIM and GIM (adapted from [16])

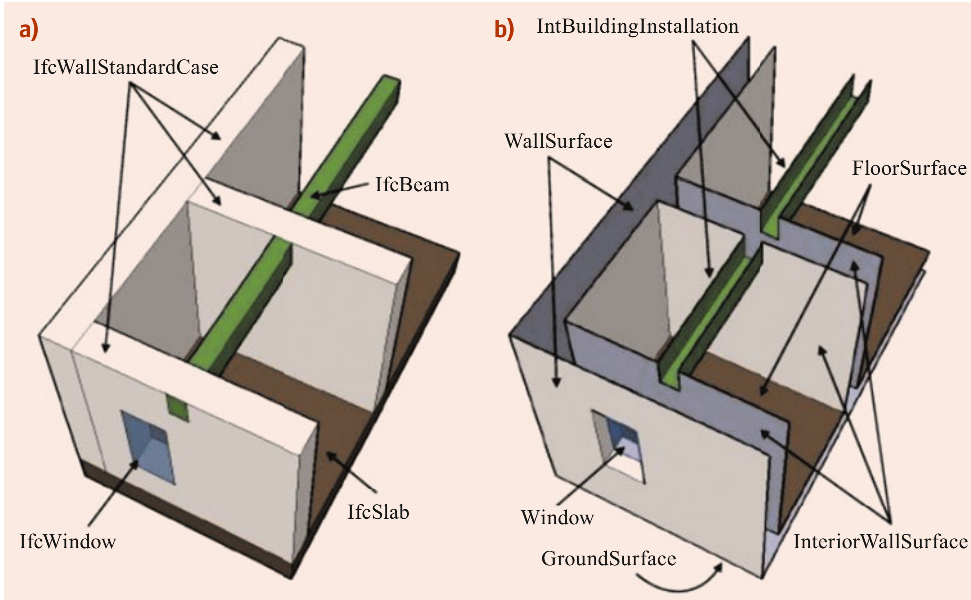


Fig. 21.14 Representation of the different modeling paradigms in IFC (**a**) and CityGML (**b**) using the example of a model excerpt; the corresponding object classes are shown (source: [18])

model emerges as a representation of reality. Accordingly, the geometry is explicitly described using node, edge, and face models (e.g., boundary representation, B-Rep). The volumetric element, e.g., a wall, is defined solely by the boundary surfaces (Fig. 21.14b).

21.4.2 Standards: IFC vs. CityGML

In accordance with the IFC standard already described in Sect. 21.2.4, open standards for data modeling and data exchange have also been developed for GIM, e.g., for city models in the OGC Standard City Geography Markup Language (CityGML) [19]. The OGC (Open Geospatial Consortium) is an international nonprofit organization committed defining open standards for the global geospatial community [20]. The differences between the object representations of BIM and GIM are reflected in different standards and make data exchange between systems difficult. Possible solutions are shown in Sect. 21.5.

21.4.3 Level of Development vs. Level of Detail, Accuracy

Since the model gradually becomes more detailed during the various planning phases in BIM, the level of development (LOD_{BIM}) have been established to match the various planning phases (Sect. 21.2.2). The GIM, on the other hand, depicts the real world, which already exists in full detail. Thus, it is necessary for GIM to determine the real-world content that should be modeled geometrically and attributively and transferred to a GIS. The level of detail (LOD_{GIM}) system has been established for 3-D city models. The LOD_{GIM} of city modeling follows the principle of generalization, which means that at a low LOD_{GIM}, smaller or less significant building parts or structures are not modeled or simplified. The LOD_{GIM} principle for 3-D city models when using the standard CityGML is shown in [21] (Fig. 21.15). The current CityGML standard, version 3.0 [22], defines four LOD_{GIM} (Fig. 21.16). This classification is also used in the German state survey (Fig. 21.17; [23]).

In GIM, the concept of accuracy primarily includes the surveying accuracy (see, for example, ISO 5725 or DIN 18710). When deriving a BIM model from surveying, the accuracy of geometric modeling must also be considered (Sect. 21.2.2).

21.4.4 Content, Scale Range, and Coordinate Reference System

The BIM method was initially developed and introduced for the construction of individual buildings. Planning extends from the basic structure of the building to details such as individual elements of technical building equipment or fixing anchors. From the planning perspective, it is practical to refer to a local coordinate reference system with a scale of 1. Alignment (orientation) of the construction axes with a rectangular building structure also simplifies modeling and measuring.

GIM has its origins in the mapping of the real world for large-scale applications such as land, space, environmental, and network information systems. In addition to describing large parts of the Earth's surface (for example, topography

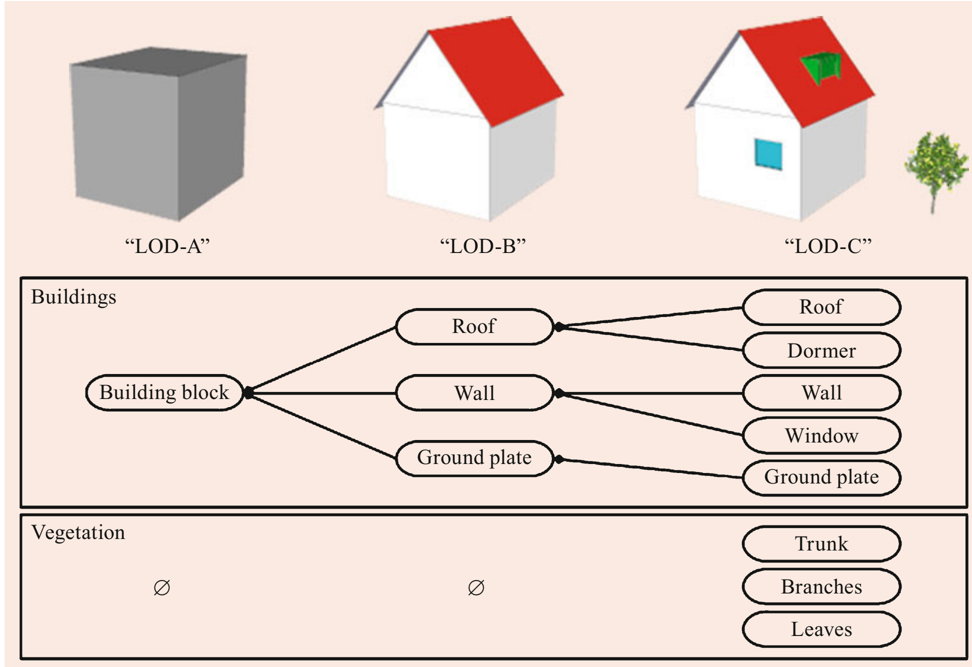


Fig. 21.15 Simplified representation of the LOD_{GIM} concept for 3-D city models and the corresponding model elements (source: [21])

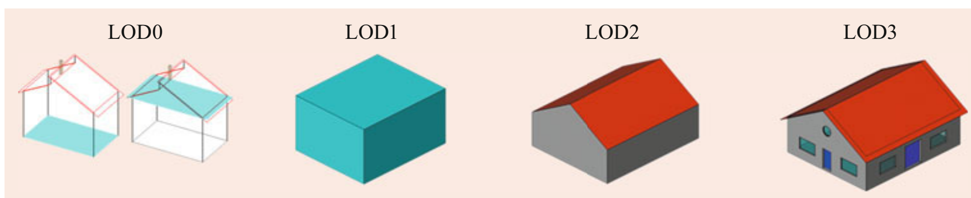


Fig. 21.16 The four LOD_{GIM} defined by CityGML (source: [22])

and terrain models), landscape and infrastructure planning (roads, etc.) are also addressed in GIM. To map the Earth's surface at this scale range, spatial coordinate reference systems (e.g., ETRS89 with the UTM map projection) are used. Length reductions caused by the curved surface of the Earth should not be disregarded when using projected coordinates (for example, UTM), which are frequently employed for official geodata (Fig. 21.18).

Mapping from the project's altitude to the reference plane of the height system used (the normal zero height (NHN) in Germany) also leads to a length correction called height reduction.

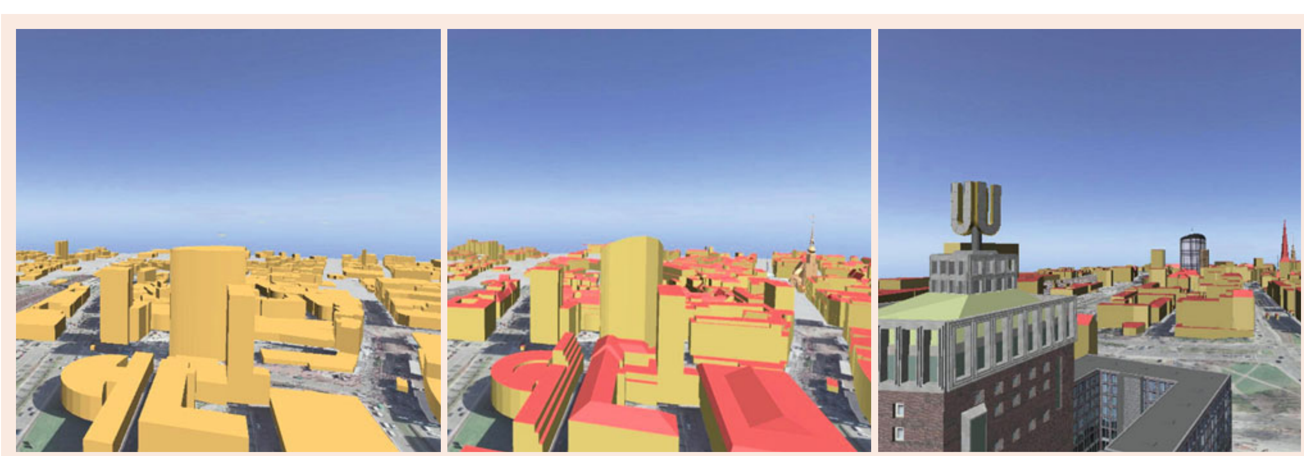


Fig. 21.17 LOD_{GIM} 1–3 for urban modeling, using GeoInfoDok 7 as an example (source: [23])

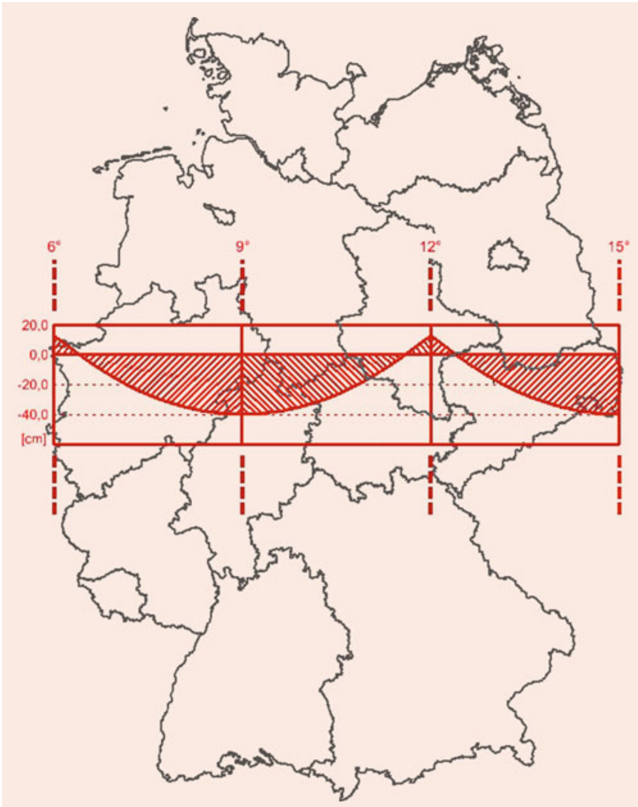


Fig. 21.18 UTM projection distortions for Germany in cm/km

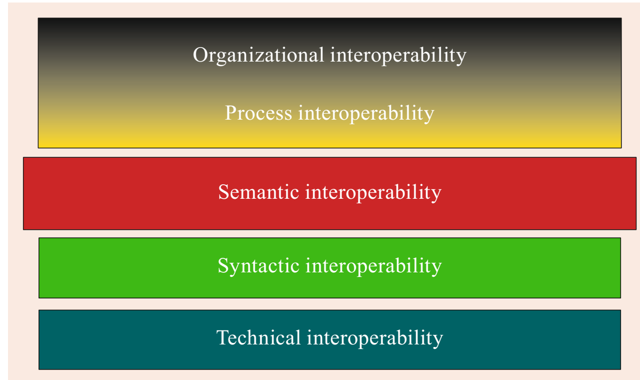


Fig. 21.19 Different levels of interoperability

21.5 Interoperability of BIM and GIM

In many cases, the planning, construction, and operation of structures require the consideration of further complementary data and information from other domains. With IFC (Sect. 21.2.4), a new exchange standard has been and is currently being created to enable collaborative work with BIM in the sense of so-called *open BIM*. Nevertheless, if existing object catalogs have also been incorporated into the standardization process for IFC, and they continue to be applied, one will not be able to manage without interfaces to data structures (models) and nomenclature of existing software, exchange formats, and object catalogs.

With the increasing use of the BIM method, in particular for large-scale construction projects such as infrastructure construction or urban neighborhoods, data from GIM are gaining in importance due to the already mentioned complementarity of BIM and GIM. For example, when planning inner-city buildings, it is of high relevance to consider how the new building planned with the BIM method integrates into the existing urban environment. Data at urban district resolution (i.e., a city model) are commonly found in GIM along with the corresponding spatial information models (e.g., CityGML) in spatial coordinate reference systems. During the planning of large-scale structures in civil engineering (for example, infrastructure constructions), information about the subsoil, the course of the terrain (the terrain model), and the topography or existing development is also required.

The different modeling paradigms must therefore be linked and intersected in order to allow, e.g., visualizations and interdependent evaluations and analyses. Thus, interoperability between the data models of the GIM and BIM must be established. In terms of information technology, interoperability can, in principle, be achieved at various levels (Fig. 21.19) (see, for example, [24, 25]), and is currently being studied intensively in both the BIM and GIM domain (e.g., [26–29]). Particular emphasis should be placed on the syntactic, semantic, and procedural interoperability of BIM and GIM.

Syntactically, BIM data differ from GIM data, e.g., in the modeling approach and associated data exchange formats used (Sect. 21.4; [17]). Semantically, the linked elements or object classes and their attributes must be assigned to each other as much as possible. In terms of process, georeferenced data must be considered in the spatial connection of different coordinate reference systems. In this case, the data must be converted or transformed into the other coordinate reference system [30].

The following part focuses on semantic interoperability between GIM and BIM, i.e., interoperability at the data model level. Various approaches can be used to establish semantic interoperability. Following *Hijazi* and *Donaubauer* [31], these are model transformation, the use of a superordinate model, or the connection of the models via a link model (Fig. 21.20).

21.5.1 Model Transformation

In model transformation, the domain-specific data of the initial model, e.g., GIM, are transformed by conversion and mapping into the data model of the target system, e.g., IFC,

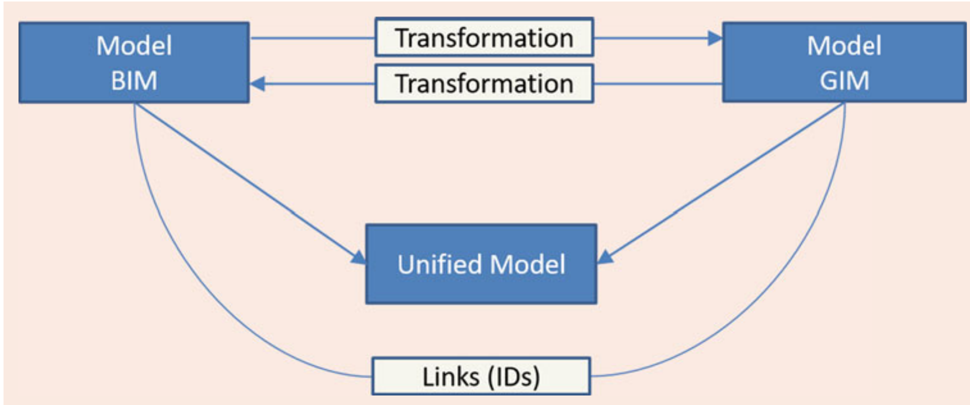


Fig. 21.20 Linking BIM and GIM (adapted from [31])

based on a suitable transformation rule. Model transformation is the most direct route to interoperability and is most commonly followed in practice using ETL (extraction, transformation, load) tools. However, the presence of different modeling levels or modeling depths is problematic as it leads to information loss or gaps [32]. In this case, mapping catalogs must be set up for the individual objects and their attributes. Also, the presence of different geometric modeling approaches (planar or volumetric) can lead to missing uniqueness and thus difficulties with geometric object creation.

21.5.2 Unified Model

In a superordinate model, the information in both worlds is fully preserved, since both models (for example, BIM and GIM) are simultaneously integrated into an overall system and thus retain their full informational content. Such a model was exemplified by *El-Mekawy* et al. [33] for IFC- and CityGML-based models and was tested using an application for building evacuation [31]. However, the creation of the parent model requires harmonization of the data models involved. In addition, a data sink and software that are supportive of both models are necessary, but these are rarely available in practice.

Standardization increasingly tries to harmonize the modeling paradigms of both worlds. For example, during the development of LandInfra/InfraGML as a potential LandXML successor for the standardized description of the landscape and physical infrastructure, OGC and buildingSmart have worked together to improve interoperability between BIM and GIM from the outset.

21.5.3 Multi-Model/Link Model

Unlike model transformation and the use of a unified model, the data in link models remain in their original data structures. There is a main system to which the data in other data structures are linked (e.g., IDs). The main system only stores links to the external systems and identifiers of the connected objects. The object information itself is aggregated case-by-case from the external model at the application level. This form of connection may be local or web-based and may occur either at the process level, e.g., via web services, or at the data level, e.g., using technologies from the semantic web [26, 34].

21.6 Dissemination of BIM

The process of introducing BIM is being driven in many countries around the world. However, it has progressed to different degrees. An overview of the distribution of BIM globally can be found in Fig. 21.21.

Singapore, Scandinavia, Great Britain, the USA, and Australia are particularly advanced [35–37]. In these countries, guidelines have been developed (indeed, there is already a second or third version of the guidelines in some cases) that mandate the use of BIM in construction projects, possibly depending on the construction volume or client [38, 39]. One example is the BS 1192 series of British directives, which now cover almost the entire objective of BIM, from the planning and construction phases (PAS 1192-2:2013) through asset management (PAS 1192-3, [40]) to safety-related information (PAS 1192-5:2015, [41]) and damage information (PAS 1192-6:2018, [42]).

Other countries are currently in the process of implementing the BIM method. An example is Germany, where, as in other countries, a phased plan for the gradual introduction of BIM in the field of transport infrastructure has been published. The German *Road Map for Digital Design and Construction* [43] has three levels. In the first two stages, experiences with the BIM method are to be collected; legal questions are to be examined; recommendations, guidelines, and patterns for assignment and processing are to be created; and requirements for uniform data structures

Fig. 21.21 Examples illustrating the international distribution of BIM (status of ∼2019)

and database concepts are to be identified. When level 3 is reached, infrastructure projects of the federal government will be planned using BIM. Also for federal buildings a BIM master plan (BMI 2021) was launched with the objective of improving the efficiency and speed of federal construction in three stages with digital planning, construction and operation [44, 45].

21.7 Summary

The digital transformation is in full swing worldwide. The AECO industry and the real estate industry are not excluded, and even have a backlog. An essential part of the digitalization of the construction industry is the adoption of the BIM method. This is made clear worldwide by statements of will and government requirements for the (obligatory) introduction of BIM, practical examples of companies in many domains involved in construction and operation, as well as international standardization efforts (e.g., those of buildingSmart International) in building construction and in infrastructure construction and operation.

BIM is increasingly becoming part of the construction, operation, and property management processes. Because it is an object-oriented modeling method, it changes the way data are handled. At least when BIM is applied at a larger scale (roads, rail, etc.), the integration of BIM and GIM becomes essential, and appropriate interoperability concepts are necessary. The challenges involved in integrating the data from BIM and GIM, such as the handling of the various coordinate reference systems or modeling paradigms, must be mastered. Standardization of the data models and processes required for collaboration are progressing rapidly.

By definition, the use of BIM is not limited to the planning phase that has been our focus so far; it can be applied throughout the entire lifecycle, particularly during the operational phase. In this phase, the BIM model, including the associated spatial information from the GIM, must become the digital twin of the real structure. Linked to this is the requirement for the model to be constantly updated to ensure that it always corresponds as closely as possible to the real structure. Only a current digital twin can be the basis for

Fig. 21.22 Augmented reality and BIM

conclusions and decisions. Classical methods as well as new technical developments such as the Internet of Things can be used to transmit up-to-date information. Augmented reality techniques, in which models or model information – possibly combined with BIM and GIM – are virtually superimposed on reality in suitable end devices, promise new possibilities in this context (Fig. 21.22).

Further Reading

1. Van Nederveen, G.A., Tolman, F.P.: Modelling multiple views on buildings. Autom. Constr. **1**(3), 215–224 (1992)
2. NIBS: National Building Information Modeling Standard Version 1.0 – Part 1: Overview, Principles, and Methodologies (2007). https://buildinginformationmanagement.files.wordpress.com/2011/06/nbimsv1_p1.pdf
3. NIBS: National BIM Standard-United States (2015). https://www.nationalbimstandard.org/files/NBIMS-US_FactSheet_2015.pdf
4. Becker, R., Falk, V., Hoenen, S., Loges, S., Stumm, S., Blankenbach, J., Brell-Cokcan, S., Hildebrandt, L., Vallée, D.: BIM – Towards the entire lifecycle. Int. J. Sustain. Dev. Plan. **13**(01), 84–95 (2018)
5. BIM Industry Working Group: A Report for the Government Construction Client Group – Building Information Modelling (BIM) Working Party Strategy Paper (2011). https://www.cdbb.cam.ac.uk/Resources/ResoucePublications/BISBIMstrategyReport.pdf
6. NATSPEC: BIM and LOD – Building Information Modeling and Level of Development. Construction Information Systems Limited, Sydney (2013). ABN 20 117 574 606
7. BIMFORUM: Level of Development Specification (2016). http://bimforum.org/lod/
8. USIBD: Level Of Accuracy (LOA) Specification Version 2.0 (2016). http://www.usibd.org/product/level-accuracy-loa-specification-version-2-0
9. El-Mekawy, M.: Integrating BIM and GIS for 3D City Modelling – The Case of IFC and CityGML. KTH Stockholm, Stockholm (2010). http://www.diva-portal.org/smash/get/diva2:390653/FULLTEXT02.pdf
10. Egger, M., Hausknecht, K., Liebich, T., Przyblo, J.: BIM-Leitfaden für Deutschland – Information und Ratgeber – Endbericht (2013). http://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-leitfaden-deu.html
11. BCF: BIM Collaboration Format (2019). http://www.buildingsmart-tech.org/specifications/bcf-releases
12. InfraRoom: Infrastructure IFC Extension Program 2017–2019 (2017). https://drive.google.com/file/d/0Bz3V7tRyy2MnWFZFNEVGazNnRzA/view
13. IFC: IFC overview (2021). https://www.buildingsmart.org/standards/calls-for-participation/ifcroad/
14. https://technical.buildingsmart.org/standards/ifc/ifc-formats/
15. PAS 1192-2:2013: Specification for Information Management for the Capital/Delivery Phase of Construction Projects Using Building Information Modelling. BSI, London (2013)
16. Hutsell, S., Bush, L.: Integrated Data Capture, BIM, CIM, GIS, and CAD-Owner and industry perspectives on products, processes and policies for informed decision making. In: SPAR 3D Expo Conf. (2016)
17. Becker, R., Kaden, R., Blankenbach, J.: Building Information Modeling (BIM) – Neue Perspektiven für Geodäten [Building Information Modeling (BIM) – New perspectives for Geodesists]. gis.Business **5**, 50–57 (2017)
18. Nagel, C., Stadler, A., Kolbe, T.H.: Conceptual requirements for the automatic reconstruction of building information models from uninterpreted 3D models. In: Proc. Acad. Track Geoweb – 3D Cityscapes Conf. Vancouver ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. (2009)
19. CityGML (2018). https://www.ogc.org/standards/citygml
20. OGC: Open Geospatial Consortium (2019). http://www.opengeospatial.org/
21. Biljecki, F., Ledoux, H., Stoter, J., Zhao, J.: Formalisation of the level of detail in 3D city modelling. Comput. Environ. Urban Syst. **48**, 1–15 (2014)
22. OGC: OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard. Version 3.0 (2021)
23. AdV: Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens (GeoInfoDok), v. 7.0.3 (2018)
24. Kubicek, H., Cimander, R., Scholl, H.J.: Organizational Interoperability in E-Government. Springer, Berlin, Heidelberg (2011)
25. EIF: New European Interoperability Framework (2017). https://ec.europa.eu/isa2/eif_en
26. Liu, X., Wang, X., Wright, G., Cheng, J.C.P., Li, X., Liu, R.: A state-of-the-art review on the integration of building information modeling (BIM) and geographic information system (GIS). ISPRS Int. J. Geoinform. **6**, 53 (2017)
27. Zhu, J., Wright, G., Wang, J., Wang, X.: A critical review of the integration of geographic information system and building information modelling at the data level. ISPRS Int. J. Geoinform. **7**, 66 (2018)
28. Sani, M.J., Rahman, A.A.: GIS and BIM integration at data level: A review. In: Int. Conf. Geomat. Geospat. Technol. (GGT 2018), Kuala Lumpur The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-4/W9. (2018)
29. Noardo, F., Arroyo Ohori, K., Biljecki, F., Krijnen, T., Ellul, C., Harrie, L., Stoter, J.: GeoBIM benchmark 2019: Design and initial results. In: ISPRS Geospat. Week Enschede The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-2/W13. (2019)
30. Becker, R., Clemen, C., Wunderlich, T.: BIM in der Ingenieurvermessung. In: Kaden, R., Clemen, C., Seuß, R., Blankenbach, J., Becker, R., Eichhorn, A., Donaubauer, A., Kolbe, T.H., Gruber, U. (eds.) Leitfaden Geodäsie und BIM. Version 3.0, pp. 84–96. DVW e.V., GIS e.V., Bühl, München (2021). https://dvw.de/veroeffentlichungen/standpunkte/4086-leitfaden-geodaesie-und-bim-version-2020-2
31. Hijazi, I., Donaubauer, A.: Integration of building and urban information modeling – Opportunities and integration approaches. In: Kolbe, T.H., Bill, R., Donaubauer, A. (eds.) Geoinformationssysteme 2017 – Beiträge zur 4. Münchner GI-Runde. Wichmann, Heidelberg (2017)
32. El-Mekawy, M., Östman, A., Hijazi, I.: An evaluation of IFC-CityGML unidirectional conversion. Int. J. Adv. Comput. Sci. Appl. **1**(3), 159–171 (2012)
33. El-Mekawy, M., Östman, A., Hijazi, I.: A unified building model for 3D urban GIS. ISPRS Int. J. Geoinform. **1**(2), 120–145 (2012)
34. Beetz, J., Borrmann, A.: Benefits and limitations of linked data approaches for road modeling and data exchange. Lect. Notes Comput. Sci. **10864**, 245–261 (2018)
35. McAuley, B., Hore, A., West, R.: BICP Global BIM Study – Lessons for Ireland's BIM Programme. CitA, Dublin (2017)
36. NBS: National BIM Report. NBS, RIBA, London (2016)
37. McGraw Hill Construction: The Business Value of BIM for Construction in Global Markets. McGraw Hill Construction, Bedford (2014)

38. May, I.: BIM in planen-bauen 4.0. In: Scherer, R.J., Opitz, F. (eds.) Bauinformatik – Baupraxis 2015: Leitlinien – Richtlinien – Normen – Ergebnisse aus dem VDI-Koordinierungskreis BIM. Tagungsbd. 6. Fachkonf. Bauinform. Bauprax. Dresden (2015)
39. Tulke, J.: Bauen digital – Stand und Entwicklung der BIM Einführung in Deutschland (2017)
40. PAS 1192-3:2014: Specification for Information Management for the Operational Phase of Assets Using Building Information Modelling. BSI, London (2014)
41. PAS 1192-5:2015: Specification for Security-Minded Building Information Modelling, Digital Built Environments and Smart Asset Management. BSI, London (2018)
42. PAS 1192-6:2018: Specification for Collaborative Sharing and Use of Structured Health and Safety Information using BIM. BSI, London (2018)
43. BMVI: Road Map for Digital Design and Construction (2015). https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/roadmap-stufenplan-in-engl-digitales-bauen.html
44. Schönbach, R., Aßmus, R., Klemt-Albert, K., Bergmann, M.: Entwicklung des Masterplan BIM für Bundesbauten. Bauingenieur **96**, 5 (2021)
45. BMI (2021): BIM Master Plan for Federal Buildings. https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2021/10/masterplan-bim.pdf
46. DIN 18710: Engineering survey
47. ISO 5725:1994: Accuracy (Trueness and Precision) of Measurement Methods and Results
48. ISO 12006-2:2015: Building Construction – Organization of Information About Construction Works – Part 2: Framework for Classification
49. ISO 12006-3:2007: Building Construction – Organization of Information About Construction Works – Part 3: Framework for Object-Oriented Information
50. ISO 16379:2013: Industry Foundation Classes (IFC) for Data Sharing in the Construction and Facility Management Industries
51. ISO 19650-1:2018: Organization of Information About Construction Works – Information Management Using Building Information Modelling – Part 1: Concepts and Principles
52. Würriehausen, F.: Semantische und Organisatorische Interoperabilität kommunaler Geodaten im Kontext von INSPIRE. Veröffentlichungen der Deutschen Geodätischen Kommission, Reihe C, vol. 749. Verlag der Bayrischen Akademie der Wissenschaften, C.H. Beck, München (2015). Dissertation

Jörg Blankenbach Jörg Blankenbach received his PhD from TU Darmstadt in 2006. Subsequently, he worked as Post-doctoral Fellow at the Geodesy Institute of TU Darmstadt. Since October 2012, he has been Head of the Geodetic Institute and Chair for Computing in Civil Engineering & GIS of RWTH Aachen University. His research interests include digital 3-D data acquisition and modeling of the built environment, geosensorics, and (indoor) positioning.

Ralf Becker Ralf Becker studied geodesy at the University of Bonn and received his degree in 1995. He prepared his PhD thesis at the Geodetic Institute of RWTH Aachen University. He has been working as Research Associate at RWTH Aachen University since 2008. His research interests are distributed GIS, BIM and computing in civil engineering.

Location-Based Services

22

Haosheng Huang

Contents

Abstract

Location-based services (LBS) have been a research field since the early 2000s. With the recent rapid advances in mobile computing technologies and the ubiquity of mobile devices, LBS are entering into many aspects of our daily life. This chapter provides an up-to-date overview of the field of LBS. Particularly, it introduces the essential components within the LBS ecosystem, i.e., mobile devices/users, positioning, communication networks, service and content provider; typical application domains of LBS; and key research areas, i.e., positioning, context modeling, and information communication and adaptation, as well as recent research trends and open research challenges currently presented in the domain of LBS.

H. Huang (✉)
Department of Geography, Ghent University
Ghent, Belgium
e-mail: haosheng.huang@ugent.be

Keywords

location-based services · mobile devices · navigation · positioning

On May 2000, US President Bill Clinton eliminated the selective availability of global positioning systems (GPS), which was an intentional degradation of GPS signals, and added 50 meters of error horizontally and 100 meters vertically. With this elimination, obtaining a user's current position with relatively high accuracy in outdoor environments became trivial. Since then, many mobile (geo-enabled) applications making use of the current position of the user have been developed. These applications are termed location-based services (LBS) or location-aware services. In recent years, LBS have become more and more popular, not only in citywide outdoor environments but also in many indoor environments, such as shopping malls and airports. Typical applications of LBS include navigation systems, tourist guides, intelligent transportation services, social networking, entertainment/gaming, mobile healthcare, etc. [1].

In general, LBS are (mobile) computing applications that provide information tailored to the current location and context of the user, very often through mobile portable devices, e.g., GPS-enabled smartphones and smart watches, and mobile networks, such as wireless telecommunication networks or WiFi networks [2, 3]. LBS can be considered as typical applications of ubiquitous computing (also known as pervasive computing or context-aware computing), which emphasize the importance of location awareness. However, LBS developers are often aware that there is more to context than location [4] and very often tailor the services according to both location and other aspects of the context of the user.

When comparing LBS with other existing geographic information systems (GIS) and web mapping applications, one can observe several key distinct characteristics of LBS. Firstly, LBS are aware of the context their users are currently in and can adapt the information and their presentation ac-

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_22

cordingly [5]. Secondly, LBS are often used in a dynamic and mobile environment and via mobile portable devices (which often have limited screen sizes, computing power, and interaction capacity). Thirdly, LBS users often have limited GIS expertise and often employ LBS to support their daily activities in space, e.g., localization, navigation/wayfinding, information/location searching, entertainment, and fitness monitoring. These distinct characteristics bring many new research questions to the GIScience research community and other cognate communities.

This chapter gives an introduction to the key components within LBS from a system perspective (Sect. 22.1), describes typical applications of LBS (Sect. 22.2), and discusses the three main tasks LBS need to address (Sect. 22.3). Finally, we outline the open research challenges that are needed to bring LBS to a higher level (Sect. 22.4).

22.1 LBS Components

For the provision of LBS, four key components commonly exist (Fig. 22.1): mobile devices, positioning technologies, communication networks, and service and content providers [6].

Some LBS applications, e.g., built-in car navigation systems, might not contain the third component, as the applications may already store all the information locally on the mobile device and process all the service requests solely on the device. In other words, sending the request over the communication network to the service and content provider is not needed [5].

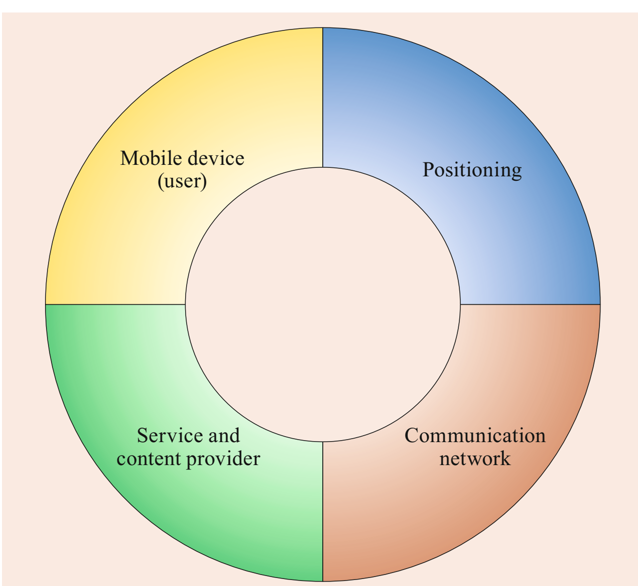


Fig. 22.1 Key components in LBS (after [6])

22.1.1 Mobile Devices

To use mobile LBS, a user has to be equipped with a mobile device, which requests and receives the needed information relevant to the user's current location. The results can be then provided in various forms on the mobile devices, such as graphical, e.g., mobile maps, augmented reality, and 3-D, verbal, haptic, and hybrid [5].

Possible mobile devices for LBS include smartphones, wearable devices like smart watches and digital glasses, haptic devices, head-mounted displays such as Microsoft Hololens 2, built-in devices like built-in car navigation systems, and so on. Among them, smartphones are still the most common mobile client for LBS. Different types of mobile devices often have their own characteristics, e.g., screen sizes, computational power, and sensors embedded; interaction modalities, e.g., touch, voice, gesture, gaze based; and technical constraints. Therefore, it is important that the LBS applications are able to adapt to these devices.

In recent years, more and more people have been carrying different mobile devices, e.g., smartphones and smart watches, at the same time. Therefore, studies have started to explore cross-device interaction in LBS [7], making use of the strengths of different devices. For example, for navigation applications, smartphones can be used to provide overview information at the beginning of the navigation process, while detailed turn-by-turn instructions can be communicated on smart watches at each decision point.

In the future, with autonomous cars and smart environments, we can expect that LBS will be accessed directly via embedded devices/display in the environment or in the car. Meanwhile, more "natural" and nonintrusive user interfaces will be developed for LBS.

22.1.2 Positioning Technologies

As the name suggests, LBS need a positioning component to determine the current location of the user. For outdoor environments, positioning via GNSS (Global Navigation Satellite Systems) or cellular networks can be used, with GNSS being the most common. GNSS, particularly the US GPS, and similar systems, such as Russia's GLONASS, EU's Galileo, and China's BeiDou, is often used and even combined to provide accurate and timely positioning information. Currently, almost all existing smartphones have built-in GNSS supports.

Since GNSS cannot provide reliable positioning accuracy within buildings, other positioning technologies are often employed for indoor environments [8], e.g., WiFi [9], RFID (radio-frequency identification) [10], Bluetooth like Apple's iBeacon, IMU (inertial measurement unit) sensors such as accelerometer and gyroscope [11], and even LED (light-emitting diode) lighting systems. Among them, WiFi-based fingerprinting has attracted significant research interest and might represent a "state-of-the-art" in indoor positioning [12]. However, achieving accurate and reliable positioning in different indoor environments is still a challenging task and requires significant research efforts. A universal or regional solution, such as GPS for outdoor environments, is still missing [3].

22.1.3 Telecommunication Networks

Telecommunication networks, e.g., cellular and WiFi networks, transfer the service request of the user from a mobile device to the service and content provider, and send the results back to the user's device [5]. Recent years have seen significant improvement regarding cellular communication technologies, from 3G to 4G, or even 5G. The evolvement over these different generations brings faster, more secure, and mobile reliable cellular networks.

22.1.4 Service and Content Providers

The service and content provider receives and processes the user request and returns the results. Very often, it needs to access and integrate different data sources, e.g., geographic boundaries, road network data, points of interest (POI) data, and event/news data, to process the service request [5].

22.2 LBS Applications

22.2.1 Classification of LBS Applications

With the rapid advances and the ubiquity of the telecommunication technologies and mobile devices, LBS enter into many aspects of our daily life, and reach out to many application domains. In general, existing applications can be classified into the following seven groups (Fig. 22.2):

1. Emergency and disaster management
2. Navigation and transportation
3. Infotainment
4. Fitness monitoring and healthcare
5. Business
6. Location-based assistive technology
7. Tracking.

Each group consists of several subgroups. Among all the top groups, "Navigation and transportation", "Infotainment", and "Tracking" are the three most popular.

22.2.2 Example Applications and Services

This section introduces some of the main application groups shown in Fig. 22.2.

Navigation and Transportation

This group of applications consists of several subgroups on navigation systems: car parking guidance, passenger information services, real-time ridesharing, and intelligent transport services.

Navigation systems are designed to support people's wayfinding tasks in unfamiliar environments. They can be classified into car navigation and pedestrian navigation, while the latter can be further divided into outdoor and indoor pedestrian navigation.

Navigation systems, especially car navigation systems, are the most successful LBS application by now and are used by many car drivers all over the world. In recent years, car navigation systems (in-car or via smartphones) have been continuously improved, e.g., with real-time traffic information or the employment of augmented reality views.

Recent years have also seen pedestrian navigation systems entering into the mass market to provide wayfinding guidance for pedestrians, for example in urban environments or even in indoor environments such as big shopping malls, airports, or big exhibition halls. These pedestrian navigation systems mostly adopt the same design principles as car navigation systems and are still in their infancy. Research findings regarding landmark-based navigation and other spatial cognition insights should be further incorporated for this kind of systems.

Real-time ridesharing, or peer-to-peer ridesharing, is another very popular LBS application that arranges one-time shared rides on very short notice. A typical example of this is the platform provided by Uber. This kind of service generally allows a rider to request a ride from his/her current location to wherever he/she wants to go, use some navigation functions to determine a driver's route and arrange and compute the shared ride. It also provides mechanisms to establish trust and accountability between drivers and riders, as well as to instantaneously handle the payments and optimize the ride matching.

Beyond this kind of service, for supporting drivers and passengers, applications for finding free on-street parking spots, safety warning, and multimodal routing have appeared.

The increasing use of these navigation and transportation-related applications has also led to increasing availability of location-based tracking data, e.g., floating car data [13]. These data are often used to estimate real-time traffic information, which can be used with other traffic data to enable intelligent transportation services.

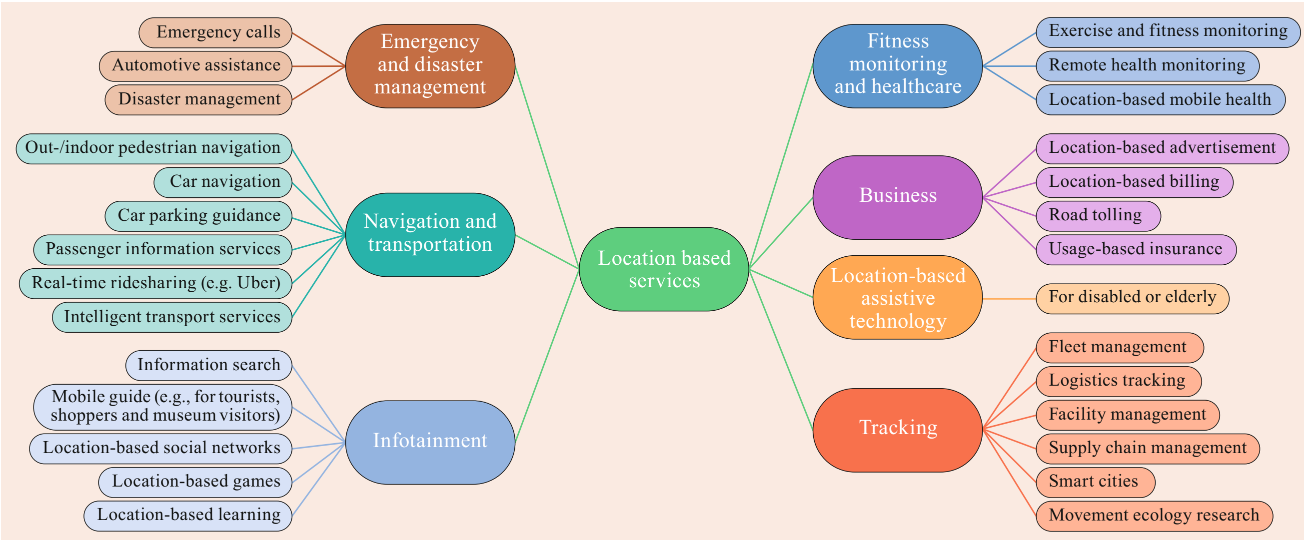


Fig. 22.2 Classification of LBS applications (after [6])

Infotainment

This group of applications consists of mobile information search services, mobile guides, e.g., for tourists, shoppers, or museum visitors, location-based social networks, location-based gaming, and location-based learning (for supporting in-field learning).

Mobile information search services, e.g., Google Maps on Android or iOS, allow users to search points of interest (POIs) around their current locations, and the research results are often visualized in a map view or list, and ranked according to distances and/or how well they match the users' preferences. These services are often combined with other LBS applications, e.g., navigation services, mobile guides or even location-based social networks.

Mobile guides are "portable, location-sensitive, and information-rich digital guides to the user's surroundings" [14]. There are three common types of mobile guides: travel and tourist guides, i.e., general mobile guides supporting tourists in various environments such as cities, airports, or recreational parks; shopping guides, e.g., supporting shoppers in big shopping malls; and museum guides, e.g., supporting museum visitors in exploring the exhibitions. They often provide functions like mobile information search, "you-are-here" maps, and navigation guidance to help users explore the environments.

In recent years, another type of LBS application, i.e., location-based social networks (LBSN), has become very popular due to the convergence of the mobile Internet, Web 2.0 technologies, and smartphones. Some prominent examples of LBSN are Foursquare, Facebook, and Instagram. They allow users to share information with friends or even the general public, connect with friends and get alerted when they are nearby, explore places and events in the real world, and receive location-based recommendations/advertisements [3, 15]. These LBSNs lead to a huge amount of location-based social media data, which have been used to generate location recommendations [16, 17], model the dynamics and semantics of cities [18], and detect and manage real-time events.

Another popular type of LBS application is on location-based gaming. Pokemon Go is probably the most known application of this type. These applications often map the real world into a virtual one, where players must move physically in real life to explore the virtual world and accomplish tasks related to the game itself [19]. The information presented on the mobile application and the tasks are adapted to the player's current location [3].

Fitness Monitoring and Healthcare

Recent years have also seen LBS being rapidly expanded to the healthcare application domain. Several subtypes of this group of applications exist [3]:

- *Outdoor exercise and fitness monitoring*, e.g., Google Fit, Running app, and Runkeeper. These applications mainly use GNSS and other sensors, e.g., accelerometers and pedometers, embedded in smartphones or smart watches to track a user's physical exercise or fitness activities and provide information such as the distance traveled, the number of steps taken, as well as the number of calories burned. To motivate and promote physical activities and healthy living styles, gamification techniques and social networking aspects are often integrated [3].
- *Remote health monitoring* applications allow doctors or caregivers to monitor the health situation of patients. For example, they can support dementia patients and their caregiver in wandering events [20]. These applications can simply make use of the GNSS tracking functions.

Sometimes, they can also integrate other sensors (wearable or on smartphones) to provide more in-depth health monitoring, e.g., to detect sudden falls of the elderly.

- *Location-based mobile health*. Recently, there has been a trend in combining LBS with other wearable sensors, e.g., for sensing heart rate, blood pressure, electrocardiogram (ECG) signals, and body temperature, for health monitoring, and providing personalized healthcare information and services [3, 21].

22.3 Core Tasks in LBS

In this section, we discuss the core tasks that each LBS application needs to address, i.e., positioning, modeling, and information communication and adaptation. The key technologies and methods of each core task will be introduced.

22.3.1 Positioning

LBS need to know the current location of the user. This section first explains the basic positioning methods, which are independent of the technologies or signals used, e.g., satellite, WiFi, RFID. We will then introduce existing positioning technologies for outdoor and indoor environments.

Basic Positioning Methods

In general, existing positioning methods can be grouped into proximity sensing, lateration, angulation, dead reckoning, and pattern matching [22].

Proximity sensing is the easiest and most widespread positioning method. The position of a target is derived from the coordinates of the base station that either receives the signals of the target, or whose signals are received by the target. This method is also known as cell of origin (CoO), which is commonly used in cellular telecommunication networks.

In lateration, the position of a target is calculated based on signals from at least n ($n \geq 3$) base stations. If $n = 3$, this method is also called trilateration. The signals are often measured by either range, i.e., the distance/travel time to a base station, or range difference, e.g., the difference of distances/travel times to two base stations. Range-based lateration is also known as time of arrival (ToA), used in GNSS like GPS. Lateration based on range difference is often called time difference of arrival (TDoA) and is often used in cellular telecommunication networks. Figure 22.3a shows the principle of range-based lateration in 2-D space; knowing one range between a target and a base station limits the possible position of a target to a circle around the base station and knowing three ranges leads to a fixed position.

In angulation, the position of a target is determined by using the signal directions (angles) between the target and a number of base stations. Angulation is also known as angle of arrival (AoA) or direction of arrival (DoA). Figure 22.3b shows the principle of angulation in 2-D space.

Dead reckoning is another basic positioning method; it estimates the current position of a target from its last known position, based on the direction of motion, and either the velocity of the target or the distance traveled.

Pattern matching compares an observation, e.g., visual images or radio signals, of the current environment to a number of prerecorded observations to estimate the current position of a target. Fingerprinting, which is often used for WiFi-based indoor positioning, is a pattern matching method based on radio signals. It contains two phases: an offline phase is executed only once to build a fingerprint database

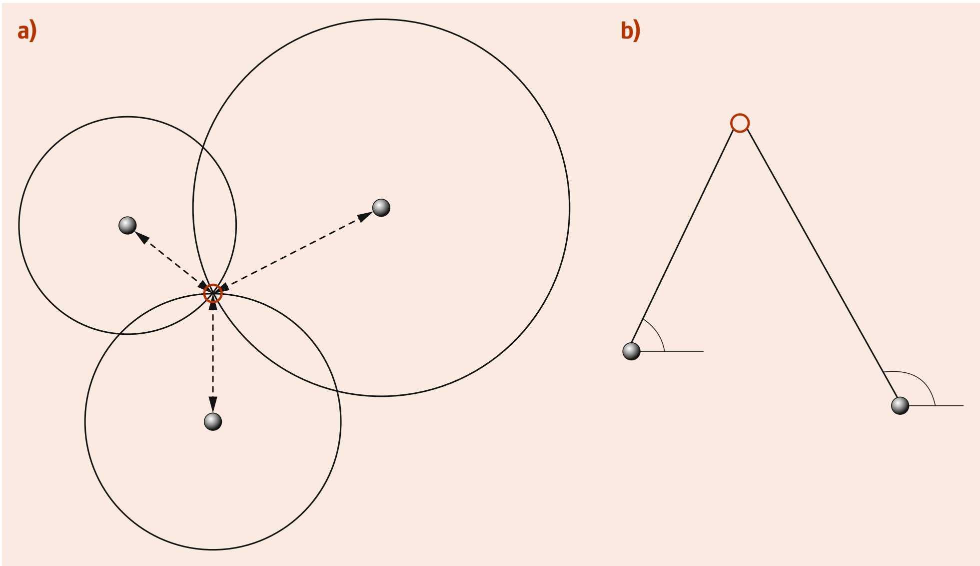


Fig. 22.3 **a** Distance-based lateration in 2-D space; **b** angulation in 2-D space, with target (*open*) and base stations (*full symbols*)

for each building. In the online phase, the target measures the radio signal strength at the current location, which is then compared with the data in the database to find the best "match".

Outdoor Positioning Technologies

Currently, the main technologies for outdoor positioning include GNSS and cellular networks.

GNSS relies on satellite signals and lateration to provide latitude/longitude and time information, anywhere on or near the Earth's surface. To ensure better positioning accuracy, the target, i.e., a mobile device such as a smartphone, needs to have an unobstructed line of sight to at least four satellites. Among all GNSS, the US GPS has the biggest market share. According to the official US GPS website (https://www.gps.gov/), GPS-enabled smartphones are typically accurate to within a 4.9 m radius under an open sky. However, their accuracy worsens near buildings, bridges, and trees. Map matching algorithms have been developed to improve the poor GPS accuracy in this kind of environment. To further improve the performance of GPS, differential GPS (to improve location accuracy) and assisted-GPS (to improve the time-to-first-fix) are also introduced. In late 2017, Broadcom introduced the world's first mass-market, dual frequency GNSS receiver for smartphones. This chip enables smartphones to provide decimeter-level positioning accuracy. Xiaomi Mi 8 is the first smartphone with this chip, and more and more smartphones with this dual-frequency GNSS chip are available for the mass market in the recent years. They potentially significantly improve current LBS applications and might open some new application fields. For example, with decimeter-level positioning accuracy, a navigation system cannot only determine which road the car is currently on but also which lane the car is using.

Cellular networks, which are mainly used for telecommunication, can be also used to locate a user. Cell of origin (CoO) and TDoA positioning methods are often employed. The location accuracy is varied, ranging from up to 100 km to 50–100 m [23]. Due to the ubiquity of GNSS chips in mobile devices, positioning based on cellular networks is becoming much less common nowadays.

Indoor Positioning Technologies

As mentioned before, GNSS positioning requires an unobstructed line of sight to at least four satellites, and cellular positioning does not provide a satisfying accuracy. Therefore, determining a user's position in the indoor environment needs to rely on other positioning technologies. Different indoor positioning technologies exist [8]: WiFi, RFID, Bluetooth, IMU sensors, and LED lighting systems. Among them, WiFi-based and Bluetooth-based technologies are the most commonly used.

WiFi-based positioning is considered to be a practical solution for indoor location tracking [5]. WiFi networks are widely available in many buildings, which provide the required hardware infrastructure without additional setup cost. In WiFi-based positioning, the fingerprint method is often employed, in which the currently measured signal strength, i.e., the fingerprint, is compared with previously measured fingerprints in the database to find a "best" match. The location accuracy of WiFi-based positioning might be up to a few meters, depending on the distribution of the WiFi access points.

Bluetooth is a communication technology that has very low power consumption and is widely available on mobile devices such as smartphones. Bluetooth-based indoor positioning often relies on Bluetooth low-energy (BLE) beacons, e.g., Apple's iBeacon, deployed in the indoor environments, and often employs the cell of origin positioning method.

22.3.2 Modeling

To ensure good usability, LBS should provide information tailored to the location, context, preferences, and needs of their users. Therefore, one of the core tasks of LBS is to effectively model these aspects. In the following, we introduce issues related to the modeling of user, location, and context.

User Modeling

Users are center to LBS applications. User modeling in LBS aims to represent users and their tasks, goals, and, thus, needs; as well as their preferences and constraints. Clustering the users in terms of their profiles, interests, and behaviors is often employed to simplify the design of LBS applications. For example, when designing LBS to guide users in a museum, the users can be classified into different groups, e.g., passionate, selective, and cursory, according to their viewing styles and interests. Based on the user groups, LBS designers can then determine what information should be provided to them and in which ways [5, 24].

The user information can be acquired by requesting the users' explicit input when using the LBS application or automatically learning from their past behaviors or interactions with the application.

Location Modeling

Location modeling in LBS mainly focuses on representing the user's current location, as well as representing geographic space and place for the LBS application [5]. As mentioned before, different positioning technologies can help to determine a user's current position, e.g., a latitude/longitude pair, or projected coordinates in a local reference system. To support various queries in LBS, the surrounding geographic space should also be effectively modeled. This can be done by integrating basic spatial data, e.g., topographic data and road network data; and other static (such as points of interest and their relevant information) and dynamic (like the location of a friend, events that are happening or will happen in a city) information. What aspects of the geographic space and place should be modeled depends on a specific LBS application.

Context Modeling

"Context is any information that can be used to characterize the situation of an entity, and an entity is a person, place or object that is considered relevant to the interaction between a user and an application" [25]. Therefore, strictly speaking, location and users are just parts of context; however, there is more to context than location and user [4, 5, 26].

An example of a classification of context factors can be found in work of *Sarjakoski and Nivala* [27]. As shown in Fig. 22.4, it consists of user, location, time (such as date, day of the week, and season), orientation/heading, navigation history, e.g., where the user has been, and what he has seen and done, purpose of the user, social and cultural aspects, physical surroundings, e.g., weather, and mobile device. This kind of classification provides some structure for identifying a potential list of context factors, while which factors to be considered and modeled depends on the specific LBS applications [3]. Different methods have been proposed to help LBS developers identify relevant context factors [28, 29].

The context information can be acquired from users via a questionnaire, smartphone sensors, e.g., GNSS and IMU sensors, wearable sensors, sensors in the surrounding environment, and online data sources, e.g., social media data, event information, and news [3]. Very often, these low-level raw data are used to derive high-level context information.

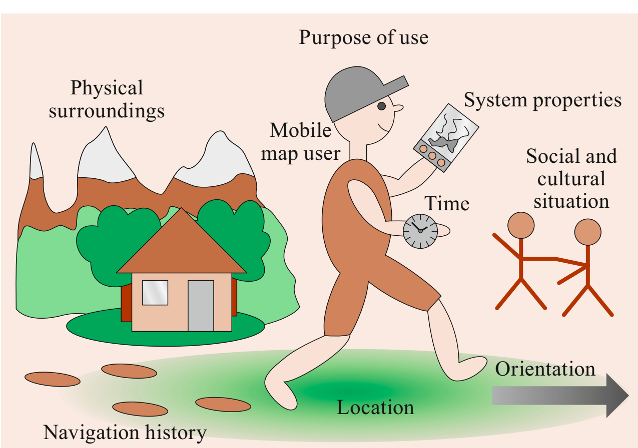


Fig. 22.4 Context factors that are potentially relevant to map-based LBS (after [6])

For the sake of clarity, we introduce the above modeling techniques related to user, location, and context separately. However, these aspects are often interlinked and should be integrated when developing LBS [5].

As can be seen from the above description, much of the data represented in these models contains personal and sensitive information, and therefore privacy concerns are often raised. This challenges LBS designers to find a balance between privacy preservation and service quality [5]. Legal regulations on this aspect, e.g., the General Data Protection Regulation in EU, should also be followed [30].

22.3.3 Information Communication and Adaptation

LBS need to communicate location-related information to their users via mobile devices to facilitate their activities in different environments. Basically, this process needs to answer the following two questions: what information is relevant to the user (and his/her location and context) and should be communicated to him/her, and in which ways, i.e., communication forms such as maps.

Communication Forms

Information in LBS can be communicated to users in various forms, such as visually, verbally, and haptically. Early LBS applications mainly communicated information to users via visual (particularly mobile maps) and auditory forms, e.g., verbal descriptions, and interfaces on smartphones [14]. In recent years, there is a trend towards more "natural" and non-intrusive user interfaces in LBS. New types of visual forms, such as augmented reality (AR), virtual reality, and 3-D are appearing [3]. Tactile interfaces, such as vibration on smartphones, have also been developed. Meanwhile, as mentioned above, in terms of interface technologies and devices, more types of mobile devices than smartphones are used in LBS, such as smart watches [31], digital glasses, head-mounted displays, and built-in displays, e.g., in cars or other environments.

Among different communication forms, mobile maps are still mostly employed in LBS [32]. One of the main design constraints of map-based LBS is the small screen size and restricted interaction capability of mobile devices, which has been one of the key research questions in the discipline of cartography since the beginning of the 2000s [33, 34] and is still a challenging task.

There are also trends of combining different communication forms, e.g., maps and voices, or even multiple mobile devices, e.g., smartphones and smart watches, concurrently to provide a better user experience in LBS [3].

Context-Aware Adaptation

As mentioned before, the contents, i.e., the relevant information, and how they are communicated to the LBS users should be adapted to the users and their locations and contexts. Therefore, context-aware adaptation plays a key role in LBS. *Grifoni* et al. [35] categorized techniques used in context-aware adaptation into four types: similarity-based reasoning, collaborative filtering, machine learning, and rule-based reasoning. Similarity-based reasoning compares a user's profile, locations, and contexts with candidate contents, e.g., points of interest for the user to visit in the mobile guide, to find a best "match", i.e., with high "similarity". Collaborative filtering selects the contents or services for the LBS users according to what other similar users in similar contexts prefer; in other words, providing services like "in similar contexts, other users similar to you often ..." [17]. Machine learning based techniques use data about a user's previous preferences and activities to first learn a predictive model and then use it to predict the future contents/services to be recommended to the user. Rule-based techniques make use of heuristics or domain knowledge to define if–then rules between context information and candidate contents/services. Depending on the current context information, some if–then rules are activated, and thus corresponding contents/services are selected for the user.

22.4 Current Research Trends and Research Challenges

Several trends regarding LBS have been identified in recent years. Some remarkable ones are [5]: from mobile guides and navigation systems to more diverse applications, from outdoor-only to indoor and mixed outdoor/indoor environments, from location-aware to context-aware, from maps and audio only to more "natural" and nonintrusive user interfaces, from technology-oriented to interdisciplinary research, and the analysis of location-based big data.

To motivate further LBS research and stimulate collective efforts, the Commission on Location Based Services within the International Cartographic Association started an initiative in May 2016 to develop a cross-cutting research agenda, identifying key research questions and challenges that are essential for the further development of LBS. The research agenda was presented in work by *Huang* et al. [3]. Particularly, it highlights a series of research challenges, organized into six groups (Fig. 22.5): ubiquitous positioning, e.g., indoor positioning, multisensor system and sensor fusion; context modeling and context-awareness, e.g., indoor spatial modeling, ambient spatial intelligence, context modeling, personalization and context-aware adaptation; mobile user interface and interaction, e.g., visualization techniques for LBS, mobile human–computer interaction, augmented reality; user studies and evaluation, e.g., mobile spatial cognition, user experiences, and evaluation methodology; analysis of location-based data, e.g., geospatial big data analytics [13]; and social/behavioral implications of LBS, e.g., privacy, legal and ethical aspects.

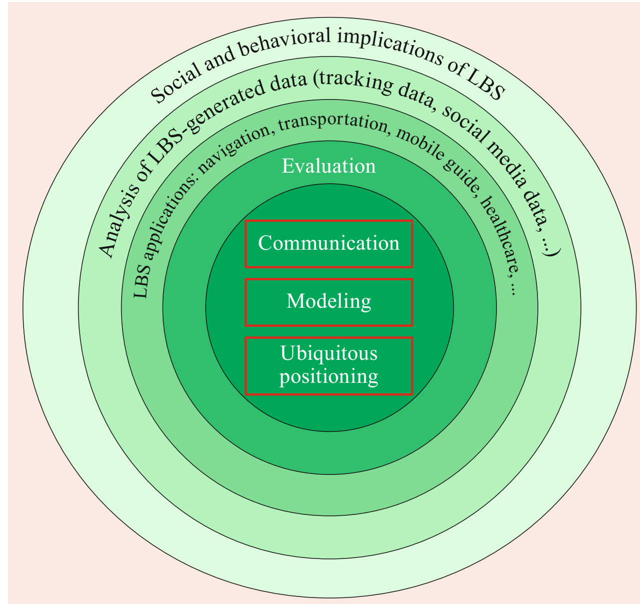


Fig. 22.5 Key research challenges in LBS

22.5 Summary

This chapter provides an overview of LBS, covering the definition, essential system components, typical applications of LBS, and the main tasks LBS need to deal with, i.e., positioning, modeling, and information communication and adaptation. Recent research trends and open research challenges are also summarized and reviewed.

With the continuously rapid evolvement of the communication technology and mobile devices, and the ubiquity of mobile devices, we can expect that significant interests from research and development will continue to be invested for LBS. One can foresee that LBS will become much smarter and enter into many more aspects of our daily life in the near future.

References

1. Gartner, G., Huang, H. (eds.): Progress in Location-Based Services 2016. Springer, Cham (2017)
2. Raper, J., Gartner, G., Karimi, H., Rizos, C.: A critical evaluation of location based services and their potential. J. Locat. Based Serv. **1**, 5–45 (2007)

3. Huang, H., Gartner, G., Krisp, J.M., Raubal, M., de Weghe, N.V.: Location based services: Ongoing evolution and research agenda. J. Locat. Based Serv. **12**, 63–93 (2018)
4. Schmidt, A., Beigl, M., Gellersen, H.-W.: There is more to context than location. Comput. Graph. **23**, 893–901 (1999)
5. Huang, H., Gao, S.: Location-based services. In: Wilson, J.P. (ed.) The Geographic Information Science & Technology Body of Knowledge (2018) https://doi.org/10.22224/gistbok/2018.1.14
6. Steiniger, S., Neun, M., Edwardes, A.: Foundations of location based services (2006). http://www.e-cartouche.ch/content_reg/cartouche/LBSbasics/en/text/LBSbasics.pdf
7. Kubo, Y., Takada, R., Buntarou, S., Takahashi, S.: Exploring context-aware user interfaces for smartphone-smartwatch cross-device interaction. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **1**(3), 1–21 (2017). https://doi.org/10.1145/3130934
8. Chen, R. (ed.): Ubiquitous Positioning and Mobile Location-Based Services in Smart Phones. IGI Global, Hershey (2012)
9. Retscher, G., Roth, F.: Wi-Fi fingerprinting with reduced signal strength observations from long-time measurements. In: Gartner, G., Huang, H. (eds.) Progress in Location-Based Services 2016, pp. 3–25. Springer, Cham (2017)
10. Bai, Y.: Development of a WiFi and RFID based indoor location and mobility tracking system. PhD thesis. RMIT University, Melbourne (2016)
11. Lam, L.D.M., Tang, A., Grundy, J.: Heuristics-based indoor positioning systems: a systematic literature review. J. Locat. Based Serv. **10**, 178–211 (2016)
12. Davidson, P., Piche, R.: A survey of selected indoor positioning methods for smartphones. IEEE Commun. Surv. Tutorials **19**, 1347–1370 (2017)
13. Huang, H., Yao, X., Krisp, K., Jiang, B.: Analytics of location-based big data for smart cities: Opportunities, challenges, and future directions. Comput. Environ. Urban. Syst. **90**, 101712 (2021)
14. Raper, J., Gartner, G., Karimi, H., Rizos, C.: Applications of location-based services: A selected review. J. Locat. Based Serv. **1**, 89–111 (2007)
15. Zheng, Y.: Location-based social networks: Users. In: Zheng, Y., Zhou, X. (eds.) Computing with Spatial Trajectories, pp. 243–276. Springer, New York (2011) https://doi.org/10.1007/978-1-4614-1629-6_8
16. Bao, J., Zheng, Y., Wilkie, D., Mokbel, M.: Recommendations in location-based social networks: a survey. Geoinformatica **19**, 525–565 (2015)
17. Huang, H.: Context-aware location recommendation using geotagged photos in social media. ISPRS Int. J. Geoinf. **5**, 195 (2016)
18. Gao, S., Janowicz, K., Couclelis, H.: Extracting urban functional regions from points of interest and human activities on location-based social networks. Trans. GIS **21**, 446–467 (2017)
19. Maia, L.F., Nolêto, C., Lima, M., Ferreira, C., Marinho, C., Viana, W., Trinta, F.: LAGARTO: A LocAtion based Games AuthoRing TOol enhanced with augmented reality features. Entertain. Comput. **22**, 3–13 (2017)
20. Pulido Herrera, E.: Location-based technologies for supporting elderly pedestrian in "getting lost" events. Disabil. Rehabil. Assist. Technol. **12**, 315–323 (2017)
21. Adibi, S. (ed.): Mobile Health. Springer, Cham (2015)
22. Küpper, A.: Fundamentals of positioning. In: Location-Based Services, pp. 121–154. Wiley, Chichester (2005)
23. Chin, K., Huang, H., Horn, C., Kasanicky, I., Weibel, R.: Inferring fine-grained transport modes from mobile phone cellular signaling data. Comput. Environ. Urban. Syst. **77**, 101348 (2019)
24. Jiang, B., Yao, X.: Location-based services and GIS in perspective. Comput. Environ. Urban. Syst. **30**, 712–725 (2006)
25. Dey, A.K.: Understanding and using context. Pers. Ubiquitous Comput. **5**, 4–7 (2001)
26. Reichenbacher, T.: Mobile cartography—Adaptive visualisation of geographic information on mobile devices. PhD Thesis. University of Zurich, Zurich (2004)
27. Sarjakoski, L.T., Nivala, A.M.: Adaptation to context—A way to improve the usability of mobile maps. In: Meng, L., Reichenbacher, T., Zipf, A. (eds.) Map-Based Mobile Services. Springer, Berlin, Heidelberg (2005) https://doi.org/10.1007/3-540-26982-7_8
28. Kessler, C.: Context-Aware Semantics-Based Information Retrieval. IOS, Amsterdam (2010)
29. Huang, H., Gartner, G.: Using context-aware collaborative filtering for POI recommendations in mobile guides. In: Gartner, G., Ortag, F. (eds.) Advances in Location-Based Services Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Heidelberg (2012) https://doi.org/10.1007/978-3-642-24198-7_9
30. Ataei, M., Degbelo, A., Kray, C.: Privacy theory in practice: designing a user interface for managing location privacy on mobile devices. J. Locat. Based Serv. **12**, 141–178 (2018)
31. Perebner, M., Huang, H., Gartner, G.: Applying user-centered design for smartwatch-based pedestrian navigation system. J. Locat. Based Serv. **13**(3), 213–237 (2019)
32. Hirtle, S.C., Raubal, M.: Many to many mobile maps. In: Raubal, M., Mark, D.M., Frank, A.U. (eds.) Cognitive and Linguistic Aspects of Geographic Space, pp. 141–157. Springer, Berlin, Heidelberg (2013)
33. Meng, L., Zipf, A., Winter, S.: Map-Based Mobile Services: Interactivity and Usability. Springer, Heidelberg (2007)
34. Griffin, A.L., White, T., Fish, C., Tomio, B., Huang, H., Sluter, C.R., Bravo, J.V.M., Fabrikant, S.I., Bleisch, S., Yamada, M., Picanço, P.: Designing across map use contexts: a research agenda. Int. J. Cartogr. **3**(sup1), 90–114 (2017)
35. Grifoni, P., D'Ulizia, A., Ferri, F.: Context-awareness in location based services in the big data era. In: Mobile Big Data. Springer, Cham (2018)
36. Google: Google Maps (2021). http://maps.google.com
37. Wikitude: Website (2021). https://www.wikitude.com/
38. Uber: Website (2021). https://www.uber.com/
39. Pokémon GO: Website (2021). https://www.pokemongo.com/

Haosheng Huang Haosheng Huang is Professor in GIScience and Cartography at Ghent University, Belgium. He received the B.Sc. and M.Sc. degrees in computer science from South China Normal University, and the Ph.D. degree in GIScience from TU Wien, Austria, in 2013. His research interests include location based services (LBS), spatial cognition, computational mobility and activity analytics, and urban informatics. Please refer to https://users.ugent.be/~haohuang/ for more details.

22

Marine Geographic Information Systems

23

Mathias Jonas

Contents

M. Jonas (✉)
International Hydrographic Organization
Monaco, Monaco
e-mail: mathias.jonas@iho.int

Abstract

This chapter describes the nautical and hydrographic application of geographic information systems. Based on a general description of the regulatory framework of the applicable performance requirements and technical norms for electronic chart display and information devices (ECDIS), a detailed explanation of the functional specifics of modern electronic sea charts follows. Emphasis is placed on the user interface and its functions for general route planning and route monitoring as well as the specific look-ahead functionality and integration of AIS (Automatic Identification System) data as significant navigation safety enhancements. Further sections deal with electronic chart data and services for their wireless dissemination. The chapter concludes with a presentation of the universal hydrographic model S-100 (based on ISO standards) as the fundamental database for e-navigation.

Keywords

International Hydrographic Organization (IHO) · ECDIS · Electronic Navigational Charts (ENC) · Automatic Identification System (AIS) · depth information · safety contour · route planning · route monitoring · look ahead function · e-navigation · universal hydrographic data model · S-100

Navigating properly is mainly dependent on obtaining the right answers to the following questions:

- Where am I?
- What are the local conditions at this position?
- Which direction should I go in to reach my destination?

The answers to this trio of questions are derived through the appropriate combination of the correctly determined ship's position with available information about the local environment. The invention of electronic means of ensuring the

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_23

availability of both at sea was the prerequisite for the development of electronic charts for navigation at sea—generally called *electronic navigational charts* in the shipping domain.

23.1 Electronic Navigational Charts for Ship Operation at Sea

It is claimed that electronic navigational charts were the first mobile GIS to support its carrier's mobility. Their development was initiated by the wide use of long-range radionavigation onboard ships. When this technology was launched in the 1980s, the British company Racal Decca combined their receivers for the hyperbolic low-frequency radionavigation system DECCA with simple line-oriented vector graphics of ship routes and shorelines, which the user then drew point-by-point using a touch-sensitive pen on a 9″ screen, much like how a graphics tablet works today. The start of satellite-based position finding—Transit in the mid-1980s, GPS later—and the introduction of digital technology onboard vessels marked the beginning of electronic navigational chart development. The first systems were presented as early as 1986, but these systems were far from perfect. It soon turned out that digital implementation of hydrographic features and specific functions customized for an independently operating craft acting under rough environmental conditions posed a special challenge. The first trials made use of vectorized data on first-generation IBM personal computers (PCs). Those trials were promising in principle, but it turned out that a number of issues had to be solved before this technology could be used routinely:

- Provision of a regular service for chart data with standardized content, and a reference system and format with global or at least large regional coverage
- Agreement regarding the definition of the scope of useful functionality to maintain ship navigation
- Availability of rugged hardware with a high storage capacity and high processing power as well as a large display size.

23.1.1 International Maritime Organization (IMO)

Unlike land-based navigation technology, marine equipment undergoes strict supervision by national and international bodies. Because shipping is international in nature, the shipping nations of the world cooperate based on regulations they have agreed upon under the umbrella of the International Maritime Organization (IMO, a special organization of the United Nations), headquartered in London. The Maritime Safety Committee (MSC) of the IMO puts in place technical performance standards for navigational equipment and proposes regulations under which the carriage of different types of equipment is acceptable or compulsory. In the case of electronic navigational charts, the IMO was not the first proponent; as the development of the new technology appeared to be primarily a matter of making printed chart material available in digital form, the first activities took place in the field of hydrography. Aspects of sea cartography are standardized by another intergovernmental organization, the International Hydrographic Organization (IHO), based in Monaco.

In 1986, a regional body of the IHO, the North Sea Hydrographic Commission, completed a study on the consequences of the development of an electronic chart display and information system (ECDIS), which became the brand name for this technology. Among its conclusions at that time were the following:

- Specifications for standardized data content, format, and updating procedures should be developed as a matter of high priority.
- To ensure the integrity of electronic navigational charts (ENCs), production should be the responsibility of hydrographic offices; ENCs should be made available in a standardized format, and all equipment should be compatible with it.
- When official ENCs are available, ECDIS users should be required to carry them, and ECDIS manufacturers or other intermediaries should not be allowed to make any modifications to the ENC data before providing the data to shipping.

As several manufacturers had begun to develop these systems, it was of immediate importance to all concerned (hydrographic offices, mariners, national shipping authorities, and manufacturers) to have at least a first draft of the IHO and IMO guidance for both the ENC and its ECDIS display systems available. The resulting draft addressed the following issues:

- Mandatory and optional data content of the ENC and required characteristics of that database, such as the cataloging of sea areas, the density of digitization of chart data, and the reliability and worldwide compatibility of chart data and other nautical information produced
- Minimum and supplementary content of the ENC display, standards for symbols, colors, and their standardized assignment to features, scale limitations of data presentation, and appropriate compatibility with paper chart symbols as standardized in the chart specifications of the IHO
- Methods for the timely update of ENCs, and means to ensure worldwide compatibility of the correction system for ECDIS data

- Design criteria for a standard format for the digital exchange of hydrographic data between hydrographic offices and for supplying the data to users, as well as procedures and financial aspects of such data exchange and supply.

Beginning in 1987, several technical working groups supported by national projects developed specifications for the ECDIS display design, comprising chart symbols and color definitions (later named IHO S-52), and the elements of S-57, the IHO Transfer Standard for Digital Hydrographic Data. S-57 is the standard to use for the exchange of ENC data.

In parallel with the development of the IHO specifications, the IMO/IHO Harmonization Group on ECDIS developed provisional performance standards for ECDIS, which were first published in May 1989 by the IMO. An amended version of the provisional performance standards based on practical experience was prepared, and this was adopted in 1995 by IMO Resolution A.817(19). At the time, all of the ECDIS proponents envisaged a short-notice breakthrough of electronic chart technology at sea—and they were half-right: PC technology made significant steps towards robustness and performance, so affordable systems conquered the bridges of the world fleet. However, the vast majority of those systems were based on chart data digitized from paper charts by private companies. It took the national hydrographic offices much longer than expected to transform their structures and skills to the regular production of ENCs with comparable coverage to paper charts. Many years later than early expectations, ENCs and ECDIS are now due to substitute paper-chart-based navigation in full. In 2009, the IMO's MSC finally endorsed carriage requirements with ECDIS, with a target completion date for all major ship classes on international voyages of June 2018. ECDIS has now become the principal navigational means and will ultimately consign paper charts to sea navigation history.

23.2 Chart Functions

23.2.1 Navigating the Vessel

The general functions of navigation are route planning, route monitoring, and route documentation. Similar to paper charts, this includes the ability of a digital system to:

- Determine the optimal route, with navigational and economic viewpoints taken into consideration
- Ensure that the route can be safely sailed, e. g., by identifying navigational aids, marking position lines, and fixing the ship's position, course corrections, and speed.

Electronic chart systems introduce a new level of performance into maritime navigation. As it is an interactive real-time navigation system, an electronic chart system is much more than a simple device to reproduce a conventional paper nautical chart on a screen. Moreover, in contrast to a paper chart, the functions of electronic charts are not limited to presenting the included information graphically. Complex integrated navigation and information systems make it possible to present selected situational information from a complex dataset together with navigation sensor information, all to support ship navigation through route planning and monitoring in a very effective way. Therefore, they also include all basic navigational functions as well as safety-relevant planning, monitoring, and control functions. Generally, the functions of electronic charts relate to:

- Basic settings (e. g., display category, highlight danger)
- Navigational elements (e. g., own ship, variable-range marker (VRM), lines of position)
- Specific functions for route planning (e. g., creation of legs, route check)
- Specific functions for route monitoring (e. g., look ahead, past track).

23.2.2 User Interface

Navigators gain access to electronic navigational charts through the user interface, also called the human-machine interface (HMI). The HMI combines functional levels (which function should be used) with operational levels (how the function will be executed). The watch officer can quickly switch information on or off, call up the required functions in a simple way, and recognize the presented values unambiguously. Figure 23.1 shows the typical components of the user interface of an electronic chart.

23.2.3 Basic Settings of Electronic Charts

The basic settings of an electronic chart system range from the selection of the displayed ENC information and the setting of various alarms to details such as the true-to-scale representation of the ship itself. The watch officer will configure the system appropriately for their purpose. The ability to choose the amount of navigational data to be displayed for the current situation is one of the most important advantages of an electronic chart system. Navigational objects can be added to or removed from the displayed chart. This ensures that the necessary information is presented and information overflow avoided. According to the IMO Performance Standards for ECDIS, there are three display categories for system electronic navigational chart (SENC) data:

- Display base: The *display base* is the minimum set of information that must always be displayed and cannot be

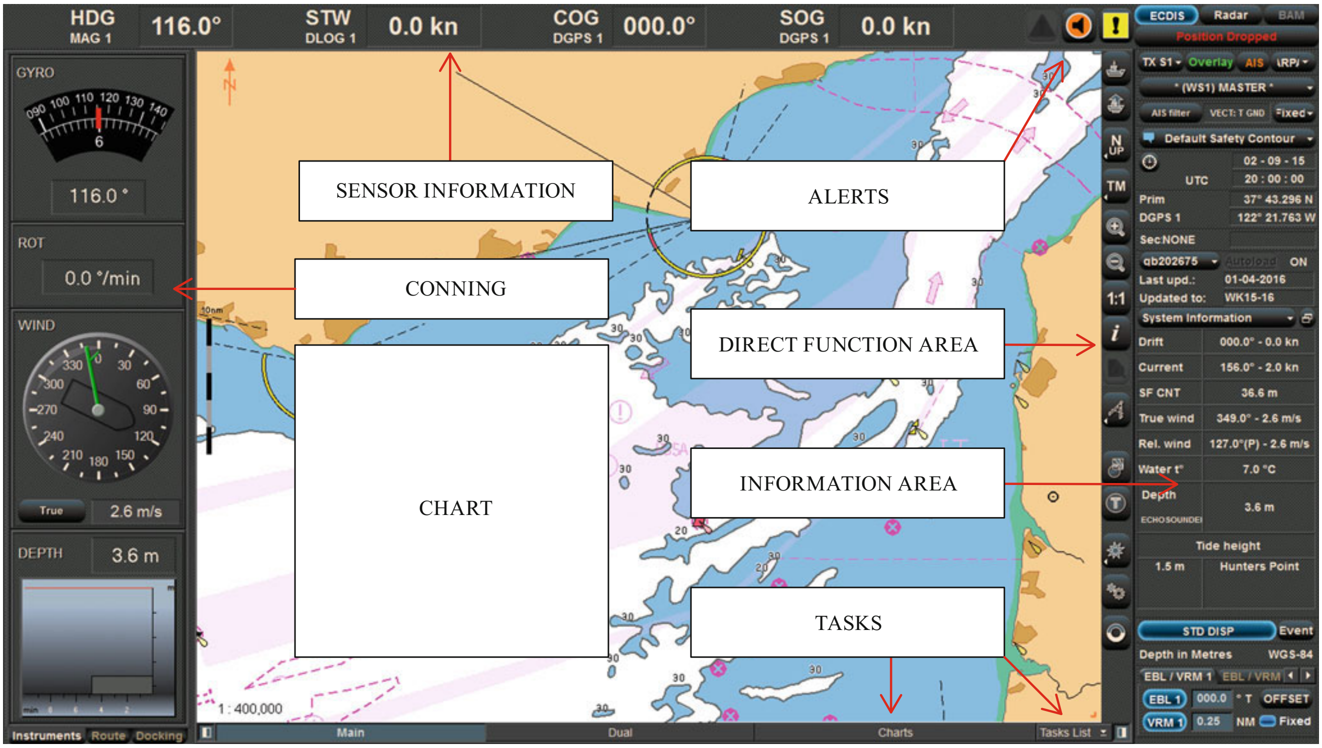


Fig. 23.1 A typical human-machine interface (HMI) of an electronic navigational chart, including a graphical *chart*, a *direct function area* for control (*right-hand side*), an *information area* below, a display of numeric sensor information (*top*), and a *function line* to monitor common navigational functions (*bottom*). Sea area: Dover Strait. (Source: Transas NS 4000 ECDIS and UKHO (data))

removed from the screen. It shows only the most important objects, e. g., coastlines, fairway, buoys, traffic separation zones, and the own ship's safety contour. In general, it is not sufficient for safe navigation but can be used to get an uncluttered overview during the appraisal and planning stage of voyage planning.

- Standard display: In normal operation during route planning and route monitoring, the *standard display* is the display mode intended to be used as the minimum set of information required for appropriate navigation. The standard display contains the display base and other important objects such as buoys and beacons, boundaries of fairways, radar-conspicuous features, and prohibited and restricted areas. If information included in the standard display is removed by the user to customize the display, this is permanently indicated. Identification of the removed information is available immediately on demand.
- All/full information: *All/full information* can be displayed on demand, as a group or individually, according to the requirements of the user and the capabilities of the system. This category includes objects additional to the *standard display*, such as spot soundings, underwater cable routes, details of navigational aids, and additional objects on land. The display can be easily overloaded with objects.

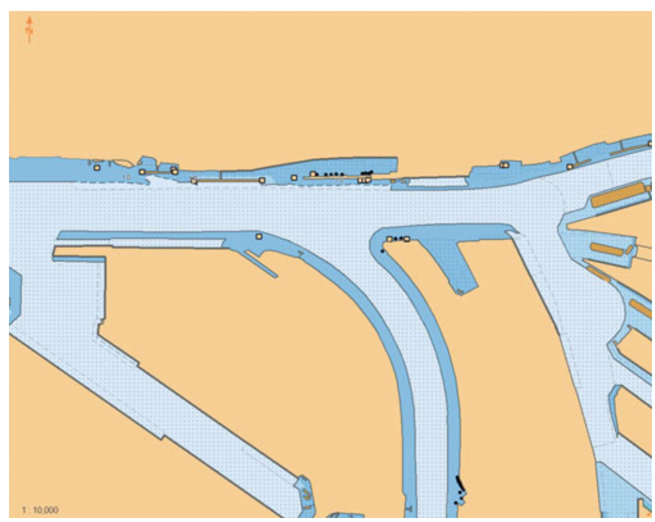


Fig. 23.2 Display category *display base*. The minimum display of ECDIS data shows, e. g., the coastline, buoys, the safety contour, and isolated dangers. Sea area: Hamburg Harbor. (Source: Transas NS 4000 ECDIS (display) and BSH (data))

To compare these three display categories, Figs. 23.2, 23.3, and 23.4 show the same sea area (Hamburg Harbor) displayed using each of those categories, along with the respective chart information.

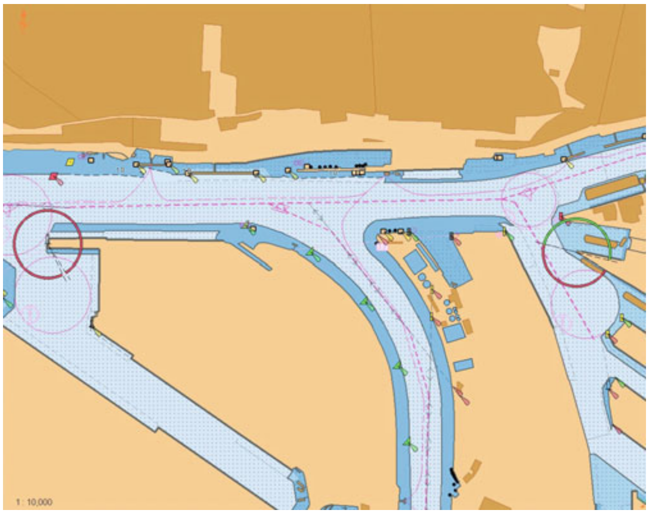


Fig. 23.3 Display category *standard display*. In addition to the display base, the display contains, for example, the 0 m contour line, fairway limits, conspicuous objects, and special areas (e. g., spoiled ground). Sea area: Hamburg Harbor. (Sources: Transas NS 4000 ECDIS (display) and BSH (data))

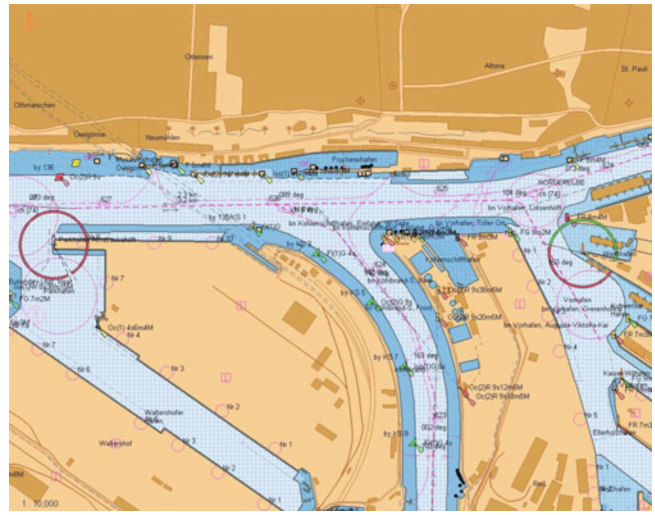


Fig. 23.4 Display category *all/full information*. This category contains additional contour lines, spot soundings, underwater cables, details of navigational aids, and additional objects on land. Sea area: Hamburg Harbor. (Source: Transas NS 4000 ECDIS (display) and BSH (data))

The user can switch from one display category to another and, if needed, add or remove objects. The active category is always indicated. At any time, it must be possible to return to the standard display with a single operator action. This is especially important when a chart is overloaded with information, or when a display base has been called up and some important information is missing.

23.2.4 Chart Selection

The immediate sea area around the own ship's current position is normally displayed during route monitoring. When the vessel is moving across an ENC-based electronic chart, the displayed chart area is adjusted automatically, using a certain internal *frame*, to show the next neighboring area. To manually call up and display any required and available sea area that does not contain the ship's position (e. g., for route planning purposes), the manufacturers use various procedures to show the chart data available in the system. Typically, a world chart (or globe) is displayed, showing the chart cells available. The user can zoom in through several steps and find the required chart—ranging from a general track overview chart to harbor and berthing details.

23.2.5 Scale, Range, and Usage of a Chart

The scale of a displayed electronic chart (both the compilation scale of the chart and the current screen-related scale) is chosen by the user. It can range from a general track overview chart (small scale) to a highly detailed harbor plan (large scale). In general, there are three different ways to describe the chart scale:

1. Natural scale (e. g., 1 : 5000 or 1 : 75,000)
2. Display range (e. g., 3, 6, 12, or 24 NM (nautical miles))
3. Chart usage (e. g., harbor, approach, or coastal).

On paper charts, the scale is given as a ratio, the natural scale (e. g., 1 : 100,000). A distance on a chart must be measured at the chart's edge with a pair of compass dividers. When using an electronic chart display, it may be desirable to express the scale as a display range in nautical miles (NM), as done when using radar. Most manufacturers provide this method. The set range (e. g., 6 NM) corresponds to the distance from the center of the chart to the top and bottom edges. Because the ship's own symbol is usually not far from the center of the chart, the display range is, in many voyage-monitoring situations, similar to the *visual* range. A change of scale from, e. g., 6 to 3 NM takes only a few seconds. This method is intuitive and also accommodates the overlaying of the radar image. Similar to what is done with radar, most systems allow the display *range* (scale) to be changed in reasonable steps.

The amount of displayed information (objects, symbols, names) depends on the display range set:

- Decreasing the range (i. e., increasing the scale) causes more detailed navigational information to be displayed automatically.
- Conversely, when displaying a larger range (i. e., a smaller scale), the image becomes more general. With a larger range, information will be cut out to avoid chart clutter.

The selected range determines, for instance, whether a certain buoy is displayed or not. The buoy may not be shown on an overview chart. However, it will appear when a certain purpose of a chart is defined or a certain range (or chart scale) is attained. The watch officer only has to change the range to get more or less detail displayed. Internally, this automatic and range-related display of details is managed by two procedures: the uploading of ENCs that are suitable for the selected range based on their level of cartographic detail and a specific individual attribution of encoded objects that defines the ranges at which the object is displayed.

23.2.6 Areas for which Special Conditions Exist

ECDIS is able to read the individual characteristics of each object explicitly—ECDIS knows the stored objects. As it is an information system, an electronic chart system that uses object-oriented vector data has the ability to recognize objects and, if necessary, to react. In particular, ECDIS can raise an indication or alarm if the ship is within a certain area or within a preset distance or time from a given area. Typical examples of such areas that are either prohibited or for which special conditions exist are traffic separation schemes (TSS), precautionary areas, and areas to be avoided. Table 23.1 contains a list of areas that ECDIS must detect and for which it must raise an indication or alarm, according to the IMO performance standards.

Table 23.1 Areas that ECDIS must detect and for which it must provide an alarm or indication. (After [1])

Specific areas (complete list)	
Traffic separation zone	User-defined area to be avoided
Inshore traffic zone	Military practice area
Restricted area	Seaplane landing area
Caution area	Submarine transit lane
Offshore production area	Anchorage area
Area to be avoided	Marine farm/aquaculture
	PSSA (particularly sensitive sea area)
Navigational hazards (examples only)	
Fixed and floating aids to navigation (also virtual AtoNs)	
Isolated dangers	
Underwater rocks at a depth less than the safety contour value	
Obstructions at a depth less than the safety contour value	
Soundings at a depth less than the safety contour value	
…	
Safety contour and constituting objects (examples only)	
Land area	
Pontoons	
Hulks	
…	

23.2.7 Detailed Background Information (*Info Box*; *Pick Report*)

Under normal circumstances, the display is free of information that is not needed. If, in other circumstances, certain information is needed, ECDIS may use the SENC database as a background information source. When interrogating a feature, the navigator may point to its position with a cursor to acquire further information. When an object is *picked*, all pertinent information about the object is shown on demand in an information window. Depending on the manufacturer, this function is called *info box* or *pick report*. In particular, the information window provides detailed information on selected objects and answers questions such as:

- What type of buoy or light is it?
- What is the meaning of this unknown object/symbol?
- What is the depth at this location?
- What is the very high frequency (VHF) channel of this reporting point?
- At what scale was the ENC compiled?

A picked position in the chart usually corresponds to quite a number (e. g., 5–20) of objects simultaneously (symbols, different overlapping areas such as a TSS, a depth area, a national territorial area, etc.) that all relate to that position. Therefore, the information displayed may be extended, and some kind of access assistance (menu, paging, register card) is offered.

23.2.8 Depth Information and Safety Contour

With ENC vector data, relevant depth information can be displayed and used for safety-related alarms. Safe (navigable) and unsafe (shallow) waters are visually indicated on the display in terms of the following items:

- A *safety contour*, e. g., the 10 m line, which can be selected and highlighted
- Several depth contours based on the depth contour intervals contained in the ENC database, e. g., 2, 5, 10, 20, 30, 40, and 200 m
- Depth areas between two depth contours, e. g., *between* 10 *and* 20 m or *less than* 10 m, which can be displayed in different colors
- Individual spot soundings, which may be indicated as dangerous (black) or safe (gray)
- Special symbols for isolated dangers (wrecks) when the depth at the location is less than the chosen *safety depth*.

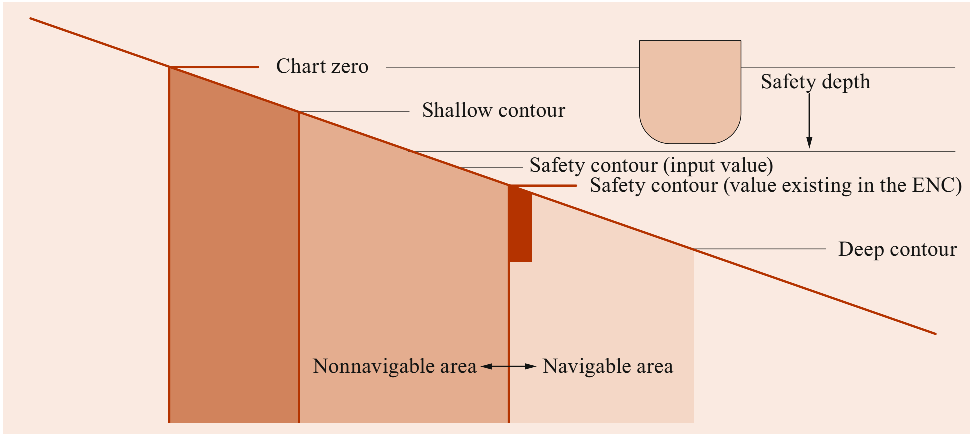


Fig. 23.5 ECDIS depth information: areas, contours, spot soundings, and the alarm-sensitive safety contour (*red arrows*). (Source: SAM Electronics and UKHO (data))

The *safety contour* is the boundary between safe and unsafe water. Based on the ship's draught, the watch officer can enter the intended safety values into the system and generate a display showing navigable waters (deep waters) and non-navigable waters (shallows). The depth areas can be discriminated by using:

- Two colors (blue and white), or
- Four colors (dark blue, light blue, gray-white, and white).

ECDIS highlights the safety contour over other contours on the display by using a bold line for it. Figure 23.5 illustrates the depth information contained in the ECDIS display: depth areas, contours, spot soundings, and the safety contour.

23.2.9 Alarms and Status Indications

ECDIS is a monitoring system offering a variety of distinguishing automatic alarms and indications to provide watch officers with the status of the electronic chart system and the navigational situation.

- *Alarms* acoustically (or acoustically and visually) announce conditions that demand particular attention—generally a danger
- *Indications* provide visual information about the condition of a system or equipment—generally a status.

A complete list of the alarms and indications officially required in ECDIS is given in Table 23.2.

Table 23.2 Required alarms and indications in ECDIS. (After [1])

Requirements	Information
Alarm	Crossing safety contour
Alarm or indication	Area with special conditions
Alarm	Deviation from route
Alarm	Positioning system failure
Alarm	Approach to a critical point
Alarm	Different geodetic datum
Alarm or indication	Malfunction of ECDIS
Indication	*Default safety contour*
Indication	*Information overscale*
Indication	*Larger-scale ENC available*
Indication	*Different reference system*
Indication	*No ENC available*
Indication	*Customized display*
Indication	*Route planning across safety contour*
Indication	*Route planning across specified area*
Indication	*Crossing a danger in route monitoring mode*
Indication	*System test failure*

23.2.10 Mode of Presentation

To display the ship's movement across an electronic chart, there are two possible modes: true and relative, both of which are well known from radar. With true motion, the own ship's symbol moves across the stationary chart based on its course and speed. This includes the need to reset the chart area from time to time. With relative motion, the own ship's symbol is stationary on the screen whereas the chart is moving in the direction opposite to the ship's course, i. e., the chart display incrementally scrolls and does not have to be reset. According to the performance standards, the true-motion mode must be available for ECDIS, while relative motion mode is an option. In an electronic chart, the two modes are more or less combined: The ship moves with its true motion within a predefined *frame* until it is a predefined distance from the chart's edge. When the own ship's symbol approaches the frame edge, the whole chart display, including the ship's symbol, is automatically reset. This is done based on the route plan and the current location of the ship to allow optimal look ahead. The frame at which the reset occurs (size, position) can be chosen by the watch officer or optimized automatically. In normal use, the reset is barely observable and the ship's motion appears, practically speaking, in true-motion mode.

Whether paper or electronic, chart information is usually shown in a north-up orientation. In certain situations, e. g., in fairways or for collision avoidance, some users prefer the course-up display, where the set course is in the upwards direction and the orientation of the chart display essentially corresponds with the view out of the bridge windows. In practice, the course-up display is only used on a bridge with much smaller ranges.

23.2.11 Navigational Tools

An electronic chart is not only a graphical display but also a *work station*. In addition to the automatic navigation functions, it provides tools to perform the conventional functions performed on paper charts (*chart work*), such as:

- Fixing the own ship's position on the chart
- Drawing bearing lines and distance rings
- Laying down the intended chart course
- Manually entering notes, etc.

What mariners do when working with paper charts can now be accomplished with equivalent electronic charts as well. Conventional functions are not sacrificed. The *tool box* of functions used on an electronic chart includes waypoints, electronic bearing lines (EBLs), variable-range markers (VRMs), distance rings, cursors, reference points, lines of position, position fixes, mariner's notes, and danger highlights. Table 23.3 lists the navigational functions required.

Table 23.3 Navigational tools, elements, and parameters of ECDIS. (After [1])

1	Own ship
1.1	Past track with time marks for primary track
1.2	Past track with time marks for secondary track
2	Vector for course and speed made good
3	Variable-range marker and/or electronic bearing line
4	Cursor
5	Event
5.1	Dead-reckoning position and time (DR)
5.2	Estimated position and time (DR)
6	Fix and time
7	Position line and time
8	Transferred position line and time
8.1	Predicted tidal stream or current vector (effective time and strength)
8.2	Actual tidal stream or current vector (effective time and strength)
9	Danger highlight
10	Clearing line
11	Planned course and speed to make good. Speed is shown in box
12	Waypoint
13	Distance to run
14	Planned position with time and date
15	Visual limits of lights arc to show rising/dipping range
16	Position and time of *wheelover*

23.2.12 Route Planning Functions

Planning a route from point A to point B is very simple on an electronic chart. The first step is to display the areas necessary for the intended route at a desired scale. The route plan is then created by entering *waypoints*, graphically rather than alphanumerically. After the entire route has been created, it should be displayed, safety checked, and finally saved for later implementation in the ECDIS as a route plan. The ECDIS performance standards require the capability to create route plans using straight and curved line segments. In the latter case, curves (of controlled or constant radius) between straight legs are created to achieve realistic behavior of a maneuvering ship. Also, great-circle routes may be created by curved line segments.

After visual inspection of the route at various scales and before departure, the selected route must be checked to determine if it can be safely navigated, i. e., if there are any known or charted dangers. The ECDIS *route-check* function checks the route for navigational risks. When performing the route check, the ECDIS compares the relevant attributes of all relevant objects along the planned route—potential risks—with data required for the safe navigation of the ship. For this purpose, the watch officer sets limits for certain safety values. The most relevant parameters are:

- Safety draught (to pass through shallow waters)
- Safety height (to pass under bridges)
- Safety distance (to pass within a horizontal tolerance)
- Planned speed and waypoint radius.

These values are specific to each individual ship and each route. They are selected according to the ship's characteristics (dimensions, draught) and the sea area characteristics (depth of available safe water, fairway dimensions). If any of these limits are exceeded during the route check, a message is produced.

23.2.13 Route Monitoring Functions

Electronic charts, particularly vector systems, contain a wide range of functions for monitoring the ship's voyage. This may range from simple visual indications to automatic monitoring alarm functions. On the ECDIS display, effective visual route monitoring can be performed by:

- Presenting the sea area at the most appropriate available scale
- Adapting the contents of the displayed chart to the actual situation (ship's draught, day or night, etc.)
- Functions such as past track and radar image overlay
- Special functions such as event marking (e. g., vessel traffic services (VTS) log-in), emergency maneuvers (e. g., person overboard), etc.

Concerning automatic alarms, effective route monitoring is essentially provided by:

- The look-ahead function for ENC objects and
- Track-related alarms (cross-track limit exceeded).

These alarms and indications serve to bring the situation to the attention of the watch officer, so that they can intervene if required.

The system can be enhanced further to become the main operating unit for automatic track control. Planned routes may include all information sufficient to operate a heading controller remotely. In combination with inputs from the position-fixing device, the gyro compass, and the log, ECDIS enables fully automatic navigation of the vessel from its starting position to its final destination, including course changes at preplanned waypoints. The electronic chart system integrates all of the functionality needed to start, watch, and—if necessary—interrupt this process according to the current situation.

23.2.14 Real-Time Ship and Environmental Data

A picture is worth a thousand words. That is why the real-time graphical display of:

- The ship's current position on an electronic chart system and
- The ship's course and speed as a vector

is of such high informational value. Nevertheless, additional alphanumeric data is also provided for the most important navigational information, for example:

- Latitude and longitude of own ship's position
- Gyro course and log speed; course and speed over ground
- Course, distance, and time to the next waypoint
- Distance and bearing of the cursor
- Cross-track distance
- Chart datum in use
- Current position fixing system (and its performance).

Other values shown may include the under-keel clearance, rate of turn, quality criteria for the position fixing system, rudder angle, setting of the machine-room telegraph, waypoint-related data, and cross-track distance, as well as environmental data such as tidal levels and the speed and direction of currents/wind. In narrow fairways or in ports, these values must be very precise. The mode of presentation varies from system to system: information bars, background pages, or a window display. Several systems let the watch officer configure the presentation of this information. As an example, such presentations may include the current status of sensors, environment, and track control; position, heading from the gyro, speed and course over ground from the differential Global Positioning System (DGPS); rate of turn; horizontal dilution of precision (HDOP) of the accessible satellites; the current; and the true and relative wind. The contents of this window may also or alternatively show very specific and detailed information, such as GPS statistics and the *best estimated position* of an integrated navigation system (INS). The contents may be reconfigured depending on the needs of the user.

23.2.15 Look Ahead Function

Even if thorough route planning—including a successful route check—has been performed, the navigator should continuously check whether the route is safe. The navigator is supported in this task by a *look ahead* function. This function aids route monitoring by continuously assessing and filtering charted objects for dangers within a user-defined area ahead. Depending on the manufacturer of the ECDIS, this function may be called *watch dog*, *safety frame*, *guard zone*, or simply *look ahead*.

The look ahead parameters of this background function can be adjusted to the prevailing navigational situation. Figure 23.6 shows a vessel in two different navigational situations. The picture on the left shows the vessel underway in open waters. The look ahead parameters are set to 12 min and 1 NM to port and starboard. The vector length of the course over ground vector is set to 12 min as well. On approaching a danger (here a patch of shallow water), the navigator will be warned well in advance. The picture on the right shows the vessel on approach to a harbor (COG (course over ground) vector length: 3 min). The look ahead has been adjusted to the situation: On entering the fairway and when proceeding to port, passing near to buoys and even land areas is a normal part of navigation. The bridge team—in this case probably the master of the vessel and the local pilot—will pay close attention to the charted sea area and their maneuvers. No announcements regarding fairway buoys or water depth are required in this situation, and the look ahead is consequently reduced to a minimum value (here 1 min ahead and 0.01 NM to port and starboard) or—if possible—switched off.

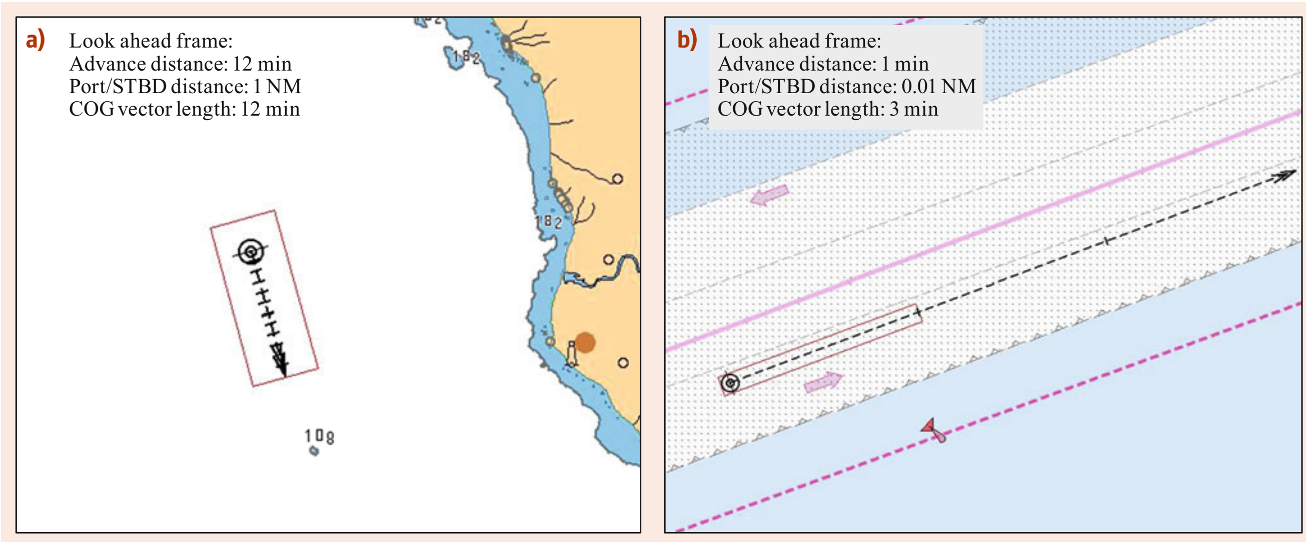


Fig. 23.6 Adjustment of the look ahead to the prevailing navigational situation. **a** Open sea: the look ahead announces dangers in due time. **b** Confined waters: look ahead announcements are reduced to a minimum

23.2.16 Predicting Own Ship's Movement

Some electronic chart systems include a very efficient function to predict the ship's movement when in a turn. The ship's position and orientation 1–3 min in the future are calculated and displayed using the prediction function and the following actual movement parameters:

- Position
- Heading
- Rate of turn
- Speed over ground.

Prediction is equivalent to intelligent dead reckoning, a traditional but still popular method to estimate the ship's position based on the known course, speed, and time elapsed since the last position fix. It provides highly useful results for short time spans into the future, in particular when performing turns in confined waters. Since only real sensor data is used for this purpose, no predetermined maneuvering characteristics of the ship are necessary. Theoretically, the prediction function is equally suitable for all kinds of ships. It is effective only in conjunction with the scaled ship symbol. Figure 23.7 demonstrates, at a large scale, the typical predicted *near future* of a ship during a course change.

23.2.17 Integration with Other Navigational Systems

The use of navigation sensors with the electronic chart enables continuous and automatic display of the actual position and movement of the ship. In an electronic chart system, nothing is more important than the real-time display of the ship's position, particularly in coastal waters, estuaries, and ports. The use of instruments that provide the ship's course and speed is a requirement. The utilization of additional sensors for, e. g., the water depth, rate of turn, wind, radar, and the Automatic Identification System (AIS) has become state of the art within the professional market. With the help of these instruments, the navigator can *drive by chart*, i. e., by monitoring the preplanned and actual tracks on the display. High availability, integrity, and accuracy of all instruments are safety-relevant prerequisites for the use of an electronic chart. The watch officer should be able to control the reliability of the input of the various sensors in use.

The main benefit of the radar overlay is the display of navigational objects and target vessels in the electronic chart. The integration of chart-related and vessel traffic information on a single display (ECDIS) provides all the information necessary for navigation and collision avoidance. In most instances, the integration of ECDIS and radar leads to improved overall situational awareness. The traffic situation is easier to comprehend, shipboard decisions are easier to make, and the workload of the watch officer is significantly reduced, since all essential information is located at one workstation. Together, these advantages increase the safety and efficiency of ship operation and makes it less expensive to operate a ship. Figure 23.8 shows a typical radar overlay with the coastline of San Francisco Bay and some target echoes. One of the inherent problems of such an overlay is the (unwanted) display of sea clutter echoes. The radar overlay on this chart has not been cleaned and is therefore *speckled* with scattered sea wave returns to a certain degree.

Automatic radar plotting aids (ARPAs) track the movements of radar targets automatically. The courses and speeds

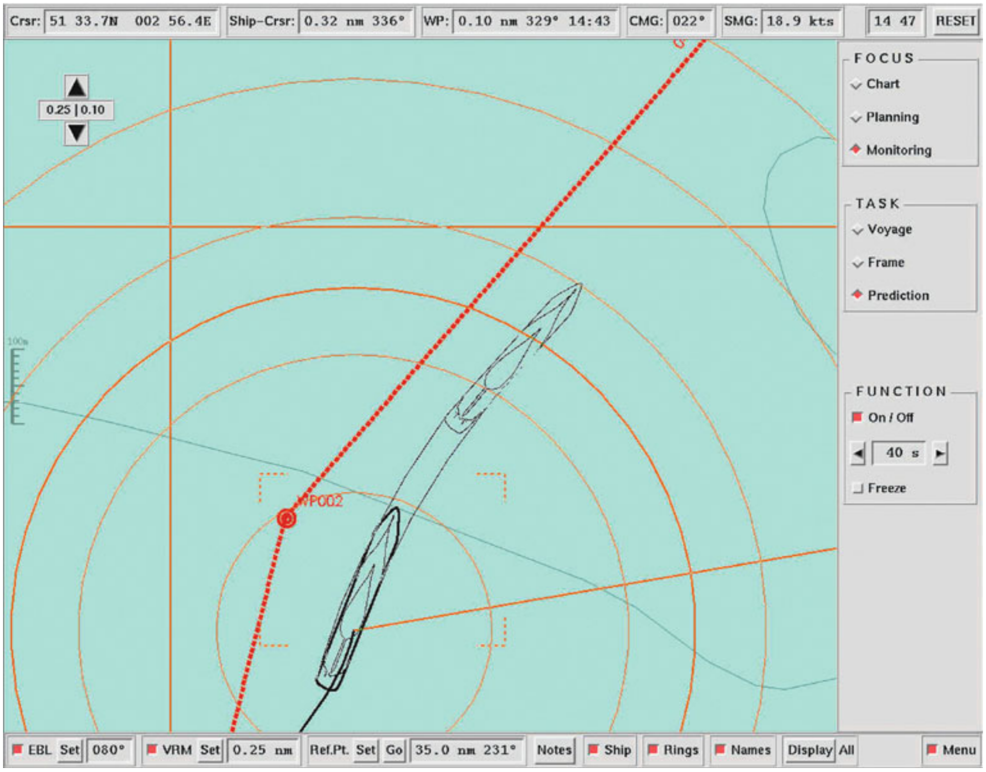


Fig. 23.7 The prediction function principle. The movement of the ship (here turning to starboard at a waypoint) a short time in the future (as predicted from sensor data only) can be displayed

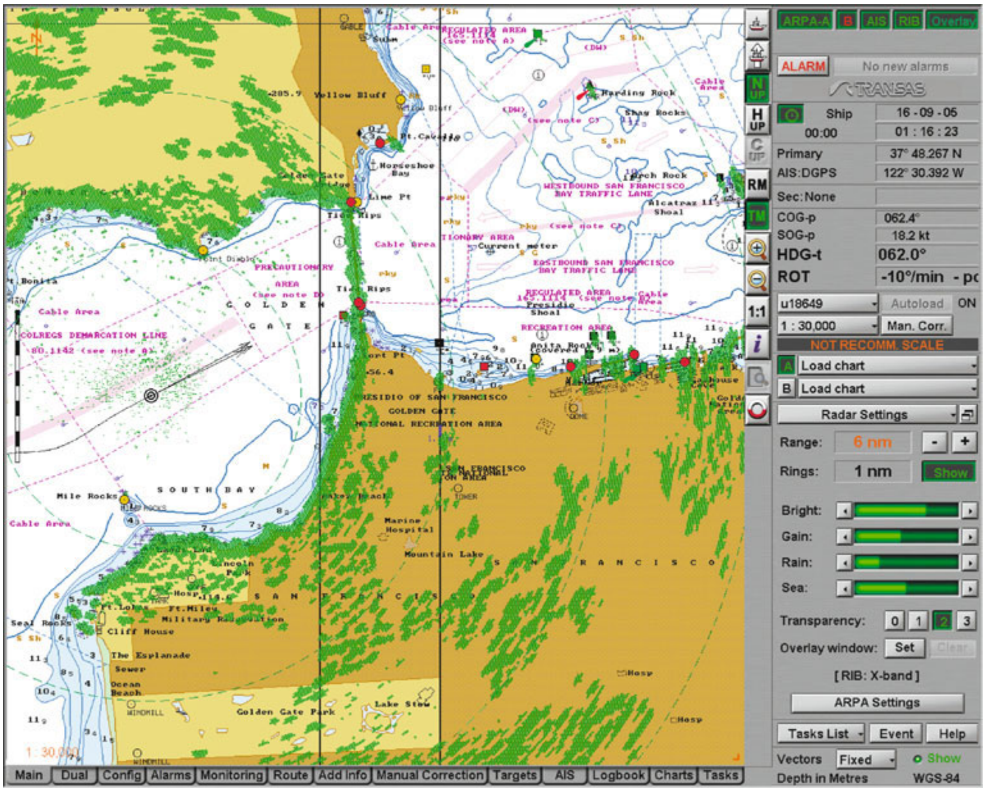


Fig. 23.8 Radar overlay on an electronic chart. The *green* radar image shows the coastline, some targets, other relevant objects (e. g., the Golden Gate Bridge), and—unfortunately—a certain amount of clutter. Range 6 NM (scale 1 : 30,000). Sea area: San Francisco Bay. (Source: TRANSAS and NOAA (data))

of ARPA targets are displayed illustratively on the radar screen using true or relative vectors, past tracks, etc. To enable this, the targets must be acquired on the radar display, either manually by the observer or automatically by the ARPA system. An example of ARPA targets displayed on an electronic chart system is shown in Fig. 23.9.

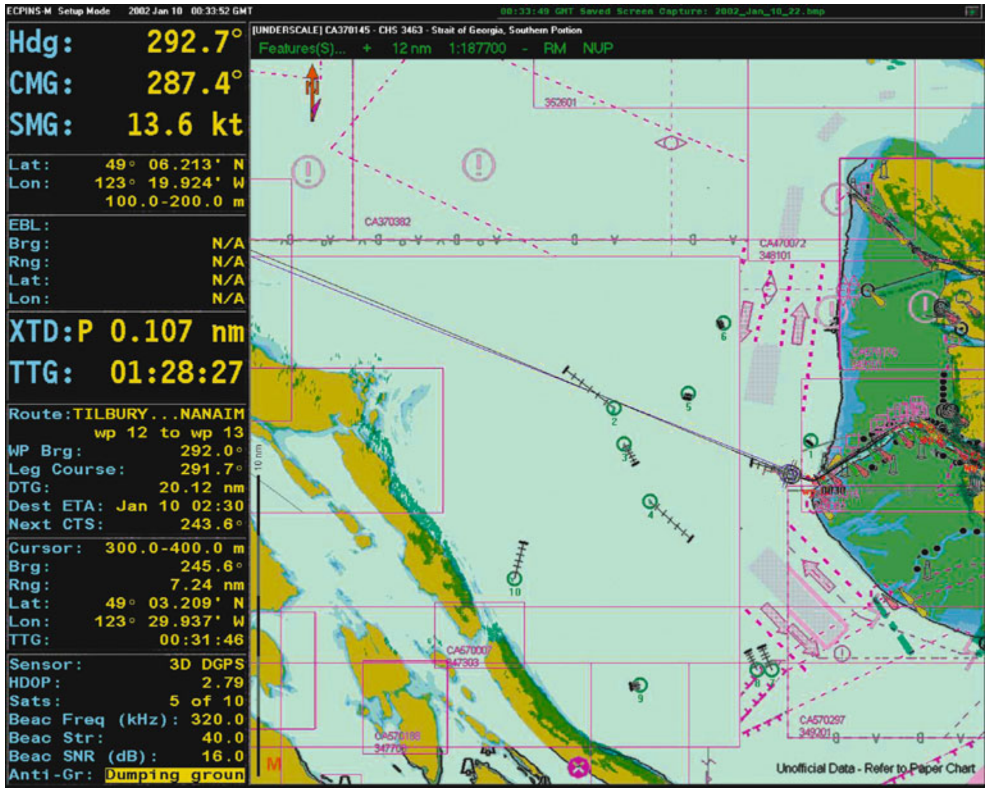


Fig. 23.9 ARPA targets in an electronic chart. The courses and speeds of all acquired targets are displayed as ARPA vectors. In addition, the complete radar video is displayed. Range 12 NM. Sea area: Strait of Georgia (Vancouver Island, Canada). (Source: Offshore Systems Ltd., Burnaby)

23.2.18 AIS Data in the Electronic Chart

Following the rapid introduction of the Automatic Identification System (AIS, a shipboard broadcast transponder system operating in the VHF maritime band), all commercial ships and many of the bigger leisure boats are now equipped with devices that continually send information about the characteristics and movements of the ship. At the same time, they receive identical information from other vessels in traffic. The self-organizing time-division multiple-access (SOTDMA) technology of the system originated in air-traffic control, and provides ship-to-ship identification, including navigational data, for a large number of participating vessels at a high broadcast rate. Each ship equipped with AIS automatically transmits its identification (e. g., international registration number, type of vessel) at short intervals as well as its current navigational data (position, course, speed, navigational status, etc.). These broadcasts mean that any AIS-equipped ship will automatically receive all the live navigational information it needs on all other vessels in traffic in the area—special manual activities are not required. AIS technology can also be used for data exchange with VTS shore-based facilities. The introduction of this identification system has broadened the field of application of the electronic chart. Upon the receipt of actual navigational data for vessels in the surrounding area, not only can these vessels be shown on the electronic chart but their movement parameters can be displayed on demand. In addition, the correlation of the navigation data broadcast by other ships with radar observations can be shown, and the echoes from vessels can be securely identified and located.

23.3 Electronic Chart Data

23.3.1 Charted Nautical Information

An electronic chart system has an enormous capacity for data storage, retrieval, and display. These stored data are the *knowledge base* of the system. They contain all geographic, hydrographic, and geophysical information for the area, the sea traffic arrangements, and the administrative regulations that are also shown on paper charts as well as described and illustrated in relevant printed nautical publications. For this reason, a distinction should be made between paper nautical charts and electronic navigational chart data. Paper nautical charts contain a fixed amount of information, while electronic chart data can contain a far greater amount of data.

23.3.2 Additional Navigational Data

In addition to the types of nautical data traditionally found in nautical publications, electronic chart systems may contain other *navigational chart* data. However, even the term *navigational* may not fully convey the fact that other types of information, normally found in numerous sources, can also be found in the electronic chart dataset. For instance, tidal constituents, explanatory texts, and even digital photographs can be stored as part of the electronic chart display system knowledge base. Temporal (i. e., time-varying) data such as ice coverage, current flow, wind direction, and meteorological events can be stored as layers of information that are periodically received and displayed as part of the electronic chart display.

23.3.3 Data Provision

All systems store the electronic chart data in some sort of database structure. However, the display of this information depends more on the data extraction, presentation mechanisms, and the utilization of the data. In the simplest sense, there are two different kinds of chart data: raster and vector. Official nautical charts (paper charts) and electronic chart data (either raster or vector) are published by the hydrographic offices of individual countries. Raster data dominated the first systems in the 1990s but now play a decreasing role, while vector chart data are about to completely take over. Currently, more than 50 countries have issued or produced electronic chart data for their territorial waters as ENCs in accordance with the official IHO vector format S-57. Worldwide coverage of official vector data, comparable to existing paper charts in coverage and resolution, has become a reality. Further improvement in ENC coverage is now primarily dependent upon new surveys or resurveys of areas where coverage is currently unsatisfactory; it is worth recalling that less than 50% of the coastal waters of the world with depths of less than 200 m are reported to be adequately surveyed.

Some private companies also produce and distribute electronic chart data. These data often provide detailed information on areas of interest for leisure boaters, i. e., smaller harbors, marinas, and shallow waters close to the shore. Their data for main waterways and adjacent areas, however, are mainly based on licensed chart data information produced by hydrographic offices. Such solutions are very popular in mobile computing on tablets and smartphones and can be found onboard leisure vessels of any size.

Studies of electronic charts by the IHO and the IMO delivered the blueprint for another interesting development: the so-called Inland ECDIS. For this adaptation of the ECDIS concept, it became evident that additional object classes, attributes, and attribute values were required to meet real-world inland navigation requirements. The European Inland ECDIS Expert Group developed a regional product specification based on IHO S-57, edition 3.1. The Central Commission for Navigation on The Rhine (CCNR) adopted it in 2006, and the Danube Commission (DC), the United Nations Economic Commission for Europe (UN-ECE), the Permanent International Association of Navigation Congresses (PIANC), and the US Army Corps of Engineers (USACE) followed thereafter. Today, large parts of major European rivers, the navigable parts of the Mississippi and Missouri rivers, and some important parts of South American rivers are already covered by inland ENCs that are in daily use on thousands of river boats.

23.3.4 Chart Corrections

Electronic chart data, like paper charts, must be current (i. e., contain the latest chart updates). One of the important labor-saving facilities provided by electronic chart systems is automatic chart updating. This is performed without the need for manual entry of corrections, in contrast to when paper charts are used. The fundamental difference between the traditional and modern methods of updating is the way in which corrections reach the ship and the method in which they are applied. More and more digital updates now reach addressees onboard through data carriers and/or remote connections, e. g., as e-mail, via satellite communication, or per mobile communication (when available within the operating area or in harbors). The time-consuming and error-prone process of manual correction has become history. The existing content of the database is updated with the new data, and the most current information is displayed immediately. However, most electronic chart systems also allow changes to be entered manually if necessary, such as when chart corrections are received via traditional printed notices to mariners.

23.4 Data Services

Private chart data producers deliver their data via a specific dealer network or install a proprietary distribution infrastructure via ground mail or the Internet. The concept of an official ECDIS data service is quite different: it was developed in conjunction with the IMO requirement to establish a globally consistent ENC updating service. Such a service has two fundamental elements:

- International consistency: All participants must work according to the same standard, i. e., the S-57 data transfer standard. All producers must ensure their ENCs exhibit consistent content and display.
- Maintenance: The vital task of providing a timely and reliable updating service throughout the world is an organizational challenge for national hydrographic offices. Added to this, updates need to reach each and every ship (whether it is in port or at sea) as quickly as possible. It would be nearly impossible for the originators—hydrographic offices—to manage this individually. Instead, they must

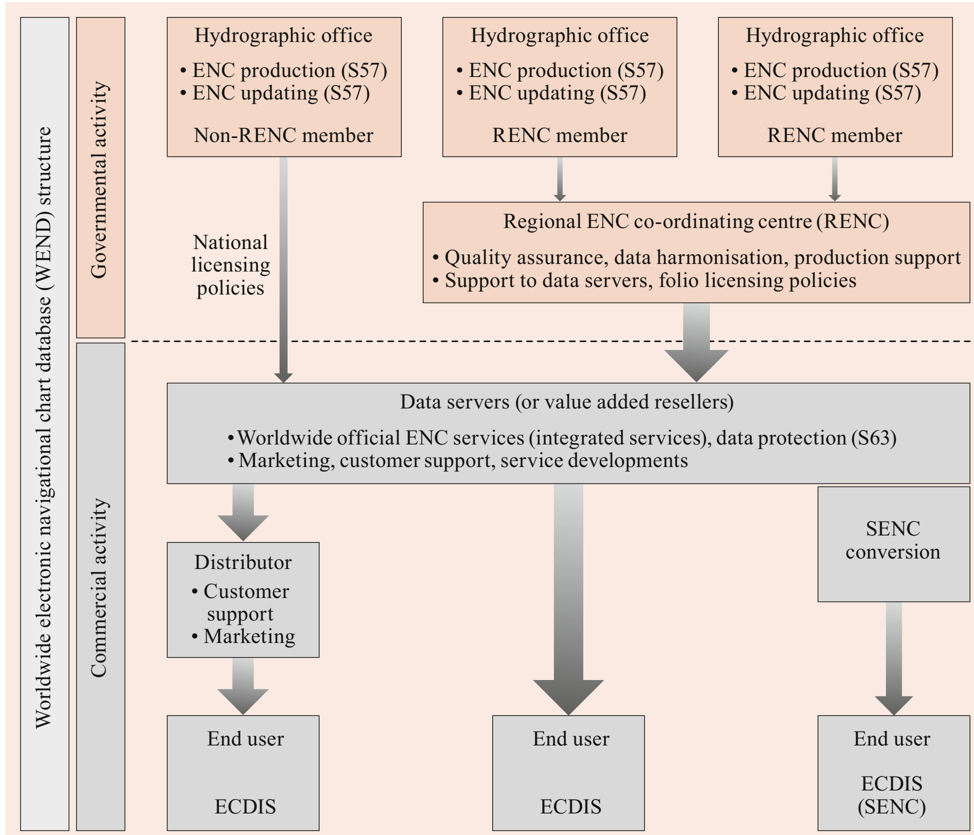


Fig. 23.10 Scheme of the WEND concept

> have effective procedures to share information and to work in close cooperation.

In an ideal world, there would be a single overarching organization to manage the official ECDIS data service. This would offer many practical advantages and achieve the desired global uniformity. However, many nations feel that ENC production is a national responsibility, and therefore a single ENC producer is not acceptable internationally. But there are practical reasons against a global solution as well, since nautical chart production needs a strong link to the local sea survey process and the compilation of associated nautical information. A reasonable compromise is to have an international arrangement for sharing the work of establishing and maintaining the provision of a worldwide ECDIS data service based upon agreed standards. Such an arrangement has been established by the IHO: the Worldwide Electronic Navigational Chart Database (WEND).

23.4.1 WEND: A Charter and an Organizational Concept

The purpose of WEND is

> to ensure a world-wide consistent level of high-quality, updated official ENCs through integrated services that support chart carriage requirements of SOLAS (International Convention for the Safety of Life at Sea) Chapter V and the requirements of the IMO Performance Standards for ECDIS. Under WEND, the term 'Integrated services' means 'a variety of end-user services where each service is selling all its ENC data, regardless of source, to the end user within a single service proposition embracing format, data protection scheme and updating mechanism, packaged in a single exchange set.

WEND consists of two components to achieve its purpose:

The first is a charter describing the principles governing the international cooperation. This includes (inter alia) the following provisions:

- By definition, the organization responsible for hydrographic survey of an area is also responsible for the ECDIS data
- The relevant ECDIS standards, especially S-57 Edition 3.1, must be observed
- The rules and responsibilities for producing and validating the data.

The second component is Regional Electronic Chart Coordinating Centers (RENCs),

> organisational entities where IHO members have established co-operation amongst each other to guarantee a world-wide consistent level of high-quality data, and for bringing about coordinated services with official ENCs and updates to them.

Hence, on behalf of its members, a RENC provides quality assurance of the ENCs and updates the members are contributing, and provides a mechanism for supporting official ECDIS data services. WEND defines the responsibilities and functions of government authorities (i. e., HOs and RENCs) with respect to developing, quality-assuring, and issuing the official ENCs and updates for use in ECDIS. It provides a framework to structure international cooperation between hydrographic offices and allows other organizations (usually commercial companies) to establish ENC services that satisfy the SOLAS carriage requirement for official, up-to-date charts. Figure 23.10 illustrates the WEND concept.

In addition to RENCs, there is another class of organizations addressed by the WEND concept. In the WEND scheme, these other organizations are termed *data servers*, but may be called *service providers or value added resellers* (VARs) by some. These entities are responsible for delivering customized end-user services, and provide the link between ENC producers and ENC users (Fig. 23.10).

So far, two RENCs have been formed, with deviating implementations of the WEND concept:

- PRIMAR Stavanger (http://www.primar.org), operated by the Norwegian Hydrographic Office
- The International Centre for ENCs (IC-ENC) (http://www.ic-enc.org), operated by the British Admiralty.

Without going into detail, one can conclude that they mainly differ in their service provision in terms of end-user contract administration. Both RENCs have expanded their member acquisition to other continents, making the regional concept superfluous. Though there would be some logic to removing the regional concept from the RENC, eventually leading to a single worldwide ENC coordinating center, there is currently no indication of such a development. Currently, the concept of the coexistence of two or possibly more hubs for ENCs looks set to remain until new technical concepts for data management and distribution such as cloud-based solutions become available at sea.

23.5 Data Display

Just as the data contents of the electronic chart system have been described as the stored knowledge of the system, the display software can be considered the necessary intelligence to display this information. To present chart information graphically on the screen, the system's software must interpret and convert the stored data to produce a chart image. In addition, it provides the operational functions needed by the navigator to use the charts effectively.

When raster data is used, the task of the display software is to reproduce the scanned paper nautical chart faithfully on the computer monitor. This is a relatively simple process whereby the pixel images of the chart are shown with the proper resolution on a monitor, and at an appropriate scale. The geographic references present in the data ensure that the display is geographically correct.

Compared with raster data, vector data is displayed in a fundamentally different and far more complicated way. Vector data contains no information on how the various colors, patterns, or symbols are to be presented. The chart image is built by first selecting the relevant chart data from the database and then assigning an appropriate color and symbol to each line, area, or object. All presentation rules to for displaying the ENC's content are contained in a separate software module—the *presentation library*. The georeferenced objects in the ENC and the proper symbolization in the presentation library are linked to each other at the moment they are called up for display. This means, of course, that the resulting image varies depending on the selected sea area, the intended display scale, and the mariner's pre-settings such as ambient light conditions and other operational conditions. The presentation library for ENCs is defined in the internationally recognized standard IHO S-52 *Specifications for Chart Content and Display Aspects of ECDIS*. The use of this presentation ensures that the official vector data look exactly the same on any ECDIS, regardless of the manufacturer. The S-52 rules were developed in close collaboration with the ENC product specifications to reproduce the whole variety of coding options equally.

In addition to ensuring the correct presentation of the colors, shapes, and symbols for the chart data, the display software provides the correct presentation of the following:

- The selected chart orientation (e. g., north-up)
- The scale indication
- The cartographic projection (i. e., Mercator)
- The own ship's position
- The range and bearing
- The symbols for radar and AIS targets
- The overlaying of a radar image.

23.6 Electronic Charts as a Component of E-Navigation

E-navigation is a core IMO initiative that aims to integrate existing/new shipboard and shore-based navigational tools into an all embracing system defined as:

> . . . the harmonized collection, integration, exchange, presentation and analysis of maritime information onboard and ashore by electronic means to enhance berth to berth navigation and related services, for safety and security at sea and protection of the marine environment.

23

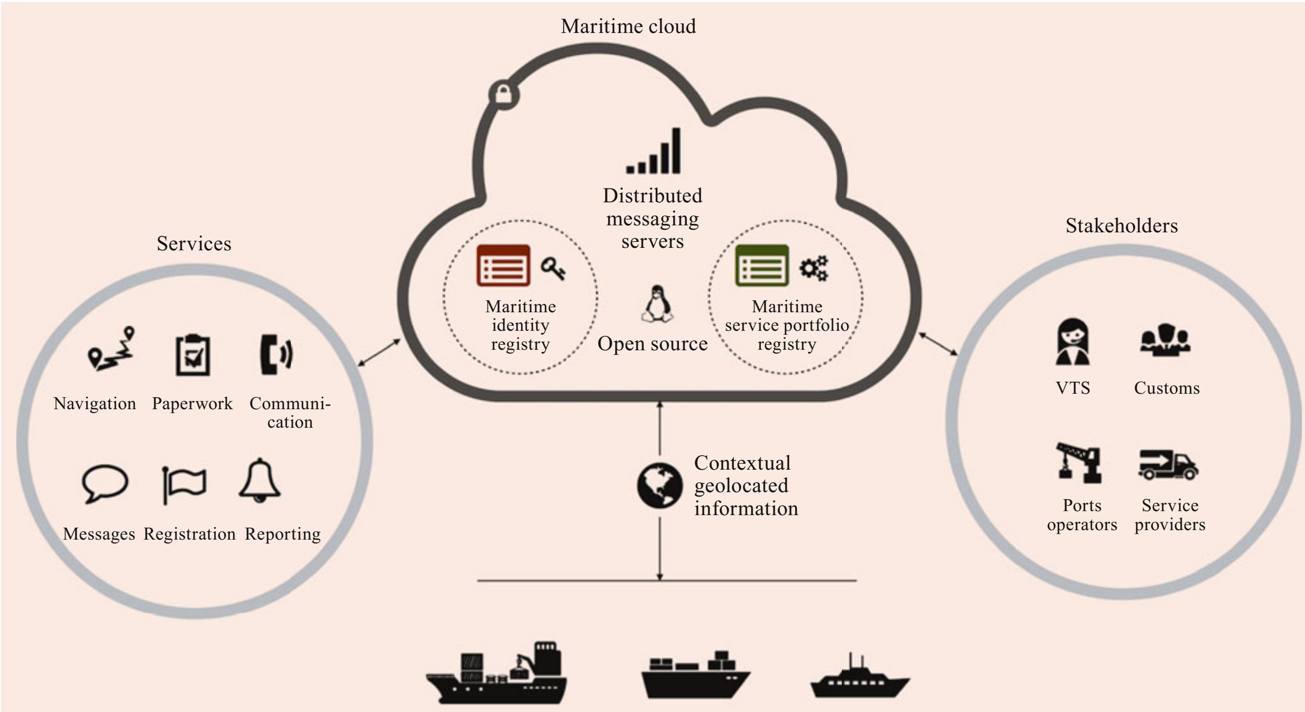


Fig. 23.11 Structure of the Maritime Cloud concept. (Maritime Cloud, 2014)

There is no doubt that chart functionalities familiar from ECDIS will be a core element of e-navigation. But e-navigation is more than an enhanced level of integration of hydrographic information within the known framework of navigational aids. E-navigation is not a sophisticated arrangement of available techniques, nor a new system, but first and foremost an operational concept of maritime navigation as a collaborative arrangement between onboard and land-based navigational facilities that continually cooperate via broadband communication channels.

Figure 23.11 presents the vision of the Maritime Cloud. According to this concept, each and every vessel will get its own individual maritime identity—a globally accessible representation of the vessel within the cloud-based digital information technology. This concept suggests that the ship's avatar will contain all sorts of static and operational information associated with the particular vessel, and this should even apply to any sort of ECDIS data fuel.

From the geoinformation point of view, improved cross-sectoral standardization of data products, improved data distribution mechanisms, modern presentation techniques such as those used in current land-based GIS systems, and the incorporation of smart algorithms for the interpretation of huge volumes of information are crucial prerequisites to achieve the vision of e-navigation. To address this, the IHO developed the concept of the Universal Hydrographic Data Model S-100 [IHO S-100, 2017].

23.6.1 The Universal Hydrographic Data Model S-100

One result of the ongoing e-navigation developments is the recommendation to IMO that the IHO data standard S-100, with its Universal Hydrographic Data Model, should be the baseline for the underlying geographic data standard. This is because of its ISO 191xx compliance and its flexibility to accommodate a wide range of different marine data. S-100 and its derivations represent the core project of the IHO's standardization branch, which aims to pave the way for e-navigation. S-100 is not a standard but a framework. It was developed to address the limitations of the IHO S-57 standard for ENCs in the first instance. But S-100 is not merely intended to provide contemporary technology for ENC coding. The S-100 concept comprises generic tools to create data product specifications, which in turn define data content that is not limited to hydrography. S-100 is founded on the Internet-based and commonly accessible IHO Geospatial Information Registry, which maintains the generation of dynamic catalogs that can be used to form data product specifications. A registry contains (by definition) a number of discrete registers and subregisters, each owned and managed by the relevant competent authority. The IHO Geospatial Information Registry contains the following component registers, each of which contains individual subregisters:

- Feature Data Dictionary Register
- Portrayal Register
- Metadata Register
- Agency Code Register.

These registers accommodate both core hydrographic content definitions and other chart-related content definitions, such as nautical publications, Inland ENCs, and marine information overlays. The IHO Geospatial Information Registry will also contain a Register for Product Specifications. The whole construct is, however, open for neighboring engineering and science domains.

Data products that are based on the registry concept benefit from two very advantageous characteristics. Firstly, their counting elements—the feature catalog and the portrayal catalog—are machine readable and can be adapted if the need arises. Since both catalogs are elements of a complete set delivered together with the data itself, the receiving device not only updates the data content but it also ingests new object definitions, rearranged attributes, and associated presentation rules. This means that for future ENCs, the chart content (including its presentation) will be updated "on the fly" without the need to adapt the kernel software of the device. The second advantage lies with the so-called interoperability of all data products derived from the S-100 registry concept. Since the basic modeling and data description elements are identical, the resulting datasets can be easily associated with each other, allowing for true integration at the end user's device. The adoption of the S-100 concept by other user domains proves the attractiveness of the concept.

S-100-Based Datasets for ECDIS

The IHO has reserved the range of the first hundred digits (i. e., S-100 to S-199) for their native data products. The flagship S-100-derived product is the future ENC format, called S-101 ENC. The S-101 ENC product specification is currently in its final development (2021) and will be released for exhaustive testing in 2022 before regular global production of S-101 ENCs commences in 2025.

S-101 ENCs are not completely different from S-57 ENCs. Aside from new arrangements for machine-readable catalogs, the set of features in S-101 is broadly identical to the set of objects and attributes already found in S-57 ENCs. Consequently, S-57 ENCs will easily be converted into S-101 ENCs one day. Backward compatibility is more difficult: since the S-101 ENC product specification offers some important improvements in the attribution and geometry of objects, backward compatibility of native S-101 ENCs with S-57 ENCs cannot be achieved. There are basically two options to solve this future problem: either upgrade all ECDISs in the field to be S-101 compatible or maintain parallel production and distribution streams for both S-57 ENCs and S-101 ENCs that cover the same area. The first option is more elegant but is less likely due to the current analog technology age IMO handling regime for software components in navigation devices. The second option can therefore be regarded as the most likely one. Admittedly, it is not yet fully clear how hydrographic service providers will serve the market with two parallel streams of ENC provision in both formats. Companies who offer digital chart data production software are currently making every effort to cover this situation, and it is confidently expected that functioning production systems will be operational in ENC production offices before the mid of this decade.

S-100-Based Data Sets for Associated Domains

The IHO has set up a nomenclature for S-100-based product specifications as part of the management of the GI Registry. The numbers S-1xx are reserved for native IHO products (Table 23.4). Blocks of numbers up to S-399 will be assigned to the main partner organizations: S-2xx numbers are assigned to IALA-domain data products (International Association of

Table 23.4 List of acknowledged S-100-based product specifications assigned to the IHO [2]

IHOS-1xx		
IHO S-101 ENC IHO S-102 Bathymetric Surface IHO S-103 Sub-surface Navigation IHO S-104 Water Level IHO S-111 Surface Currents IHO S-112 Dynamic Water Level	IHO S-122 Marine Protected Areas IHO S-123 Marine Radio Services IHO S-125 Marine Navigational Services IHO S-126 Marine Physical Environment IHO S-127 Marine Traffic Management IHO S-128 Catalogue of Catalogues	IHO S-131 Marine Harbour Infrastructure

Table 23.5 List of acknowledged S-100-based product specifications assigned to other authorities [2]

IALA S-2xx	IOC S-3xx	Others S-4xx
IALA S-201 Aids to Navigation Information IALA S-211 Port Call Message Format IALA S-212 VTS Digital Information Service IALA S-240 DGNSS Station Almanac		IEHG S-401 Inland ENC IEHG S-402 Bathymetric Contour Overlay for Inland ENC JCOMM S-411 Sea Ice Information JCOMM S-412 Weather Overlay JCOMM S-421 Route Plan

Marine Aids to Navigation and Lighthouse Authorities), and S-3xx to the IOC (Intergovernmental Oceanographic Commission). All other numbers beyond 400 will be assigned on a first-come first-served basis (Table 23.5).

23.6.2 Outlook for the Future

The replacement of the printed nautical chart by its digital equivalent was one of the first elements of the digitization of shipping. Electronic charts were the very first navigation devices to run on standard PC technology. However, when they first appeared on vessels, this revolutionary technology and new approach were not embraced by many. Indeed, the term coined for the new technology, *electronic chart*, proves this, as it suggests that an electronic chart is only the technical replication of a (printed) chart. Today, we are, as a society, acclimatized to the reality that any new tool operates electronically, so we would probably term an electronic chart system a *maritime geoinformation system* instead. ECDIS was, as a concept, cutting-edge technology in the 1990s. Comparable land-based solutions based on vector data, such as car navigation, web-based map services, and geodata-based apps, arrived years later but have quickly overtaken ECDIS in terms of technical evolution. However, electronic chart systems designed for the huge market of leisure boats have become the playground to test the user acceptance of new features such as 3D, the incorporation of crowd source information, and alternative chart symbology. The missing element at sea is, however, continuous broadband Internet access. If it becomes globally available, this will definitely be a game changer. The smartest vision in this regard is the Maritime Cloud. According to this concept, each and every ship will get its own individual maritime identity—a globally accessible representation of the ship within the cloud-based digital information technology. This concept suggests that the ship's avatar will contain all sorts of static and operational information associated with the particular vessel, and this would even apply to any sort of ECDIS data fuel.

Overall, two main trends are predicted for electronic charts. First, separate ECDIS units will disappear—they will be absorbed as a function into an integrated environment and embedded into a unified user interface concept. Second, more of the information that is presently only processed by humans will be given to machines. This will allow machines to become even more involved in the navigational process. For instance, automated analysis that takes into account the integrity of all the data could readily merge positional and dynamic information with data that are already available digitally, such as ENCs, maritime safety information, and meteorological and tidal predictions. For decisions not involving other vessels, this potentially allows some very sophisticated decision-making processes concerning navigation to be automated. If such a trend takes place, the visual presentation of chart items may generally cease to be important, with only the relevant information for the situation at hand displayed. Visual detection remains particularly important, and a demanding part of the art of human navigation is in making the right decisions when there are other vessels in the vicinity. However, whether humans will tend to remain the only detectors and users of visual information for navigation purposes that are onboard a vessel is an open question. There is currently much research into the concept of an unmanned ship; the relocation of all navigation competencies from ship to shore would see an end to the need to carry an electronic chart as an onboard device. However, without doubt, the graphical presentation of maritime geoinformation will remain the most intuitive way of understanding our environment, whatever the future brings with new technologies and automation.

References

1. IMO: Performance Standards for ECDIS, 1995 edn. (Resolution A.817(19)) and current edition (Resolution MSC.232(82)). International Maritime Organization, London (2006)
2. IHO: HSSC IRCC13 Report, Monaco (2021)

Further Reading

3. Bundesamt für Seeschifffahrt und Hydrographie: CONTIS Maps. BSH, Hamburg, Rostock (2011). http://www.bsh.de/en/Marine_uses/Industry/CONTIS_maps/index.jsp
4. Commission of the European Communities: Building a European Marine Knowledge Infrastructure: Roadmap for a European Marine Observation and Data Network. EC, Brussels (2009)
5. International Hydrographic Organization: Home Page (2010). http://www.iho.int/english/home/, Accessed 9 Feb 2010
6. Kresse, W., Fadaie, K.: ISO Standards for Geographic Information. Springer, Berlin, Heidelberg (2004)
7. Bundesamt für Seeschifffahrt und Hydrographie: Spatial Planning in the German EEZ. BSH, Hamburg, Rostock (2010). http://www.bsh.de/en/Marine_uses/Spatial_Planning_in_the_German_EEZ/index.jsp
8. Commission of the European Communities: Progress Report on the EU's Integrated Maritime Policy. EC, Brussels (2009)
9. European Commission: Legal Aspects of Maritime Spatial Planning Framework Service Contract. FISH/2006/09—LOT2. EC, Brussels (2008)
10. Krige, D.G.: A statistical approach to some basic mine evaluation problems on the Witwatersrand. J. Chem. Met. Min. Soc. S. Afr. **52**, 119–139 (1951)
11. International Hydrographic Organization: IHO Universal Hydrographic Data Model: Ed. 4.0.0. Publication S-100. (2018)

Mathias Jonas Mathias Jonas received his PhD in Nautical Engineering from the University of Rostock, Germany, in 1992. He is the Secretary-General of the International Hydrographic Organization (IHO) since 2017. Prior to this appointment he held the posts of Vice President of the Federal Maritime and Hydrographic Agency and National Hydrographer of Germany with responsibility for sea survey and sea cartography. Being a mariner, Mathias Jonas has been is involved in integrated navigation matters since the beginning of the nineties. In addition he has completed the world's first technical certification of an Electronic Chart Display and Information System in1999 and has continuously contributed to IMO and IHO standardisation activities for navigation equipment, survey and cartography since.

Geographic Information Systems in Agriculture

24

Ralf Bill, Edward Nash, Görres Grenzdörffer, and Jens Wiebensohn

Contents

Abstract

Since the beginning of the 1990s, modern agriculture and farming has changed dramatically. Agriculture has become a high-tech industry. With the possibility of locating agricultural machinery in the field using satellite positioning technologies (Global Navigation Satellite Systems, GNSS) and the increasing availability of geographic information in digital form and in increasing quality, farmers are now able to measure the spatial and temporal variability in soil, vegetation, relief, etc., within a field and to modify their operations to react to this. Farmers keep electronic field records and farm diaries, which they are able to use on site with mobile electronic devices in order to enter or retrieve information. Agricultural machinery is also being continuously developed, since due to the in-field heterogeneity many different operations (from yield mapping to plant protection) are performed on site and logged so that they may be later evaluated by computer. Due to legal regulations (IACS, cross compliance, traceability, quality management, etc.), GIS (and GeoWeb services) and information-driven crop production are becoming common tools in agriculture, which must be integrated into usual farm practices.

Keywords

precision agriculture · smart farming · farm management · land parcel identification system · information-driven plant production · site-specific crop production · farm robotics

Maps have been used for many years in agriculture, for instance, cadastral maps (Chap. 20) for the sale or leasing of farmland, or soil maps to better understand the properties of the land. These were combined with the local knowledge of the farmer and the available agricultural machinery to make field management decisions. Due to the limited amount of mapped detail and technical capabilities, decisions were invariably made at the level of whole field plots.

R. Bill (✉) · G. Grenzdörffer
Faculty for Agricultural and Environmental Sciences, Rostock University
Rostock, Germany
e-mail: ralf.bill@uni-rostock.de

G. Grenzdörffer
e-mail: goerres.grenzdoerffer@uni-rostock.de

E. Nash
DVZ Datenverarbeitungszentrum Mecklenburg-Vorpommern GmbH
Schwerin, Germany
e-mail: e.nash@dvz-mv.de

J. Wiebensohn
Premslin Farm KG
Karstädt, Germany
e-mail: geo@premslin.farm, jens.wiebensohn@uni-rostock.de

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_24

Since the start of the 1990s, modern land management has changed dramatically. Agriculture has become a high-technology industry. With the capability of locating agricultural machinery in the field using satellite positioning technologies (Chaps. 8 and 9; global navigation satellite systems, GNSS) and the increasing availability of geographic information in digital form and in increasing quality, farmers are now able to measure the spatial and temporal variability in soil, vegetation, relief, etc., within a field and to modify their operations to react to this. Farmers keep electronic field records and farm diaries, which they are able to use on site via smartphones or tablets to enter or retrieve information [1, 2].

Agricultural machinery is also being continuously developed, since due to in-field heterogeneity many different operations (from yield mapping to plant protection) are performed on site and logged so that they may be evaluated later by computer. This topic will be investigated further in Sect. 24.3. Machinery is now adorned with a plethora of application devices and onboard computers, which are now becoming standardized with the introduction of the so-called ISOBUS standard (ISO 11783). Similar to the interoperability initiatives in the geographical information (GI) community, this standardization aims to support common, lossless, continuous use of the large quantity of available data along the whole agricultural value chain.

In the European Union (EU), agriculture is tightly legally regulated, but also state supported. Applications for subsidies have, for a number of years, been performed digitally, and since 2005 directly on the basis of spatial information. This has led to the fact that, in the EU countries, agricultural parcels are fully digitized in geographical information systems (GIS), which will be further discussed in Sect. 24.2. Such regulations, and the volume of requirements with which a farmer is confronted, lead to information-driven agriculture. This requires the cooperation of all actors (farmers, various contractors, machinery syndicates, seed producers, and buyers from traders through to consumers) along the whole value-added chain and a continuous information flow and decision process.

A multitude of providers of specific geoinformation (e.g., official bodies, geoinformation brokers) are available to farmers; there are also increasing numbers of information and decision-support systems for agriculture.

With these few examples, the driving forces and the increased pressure for GIS use in agriculture at a high technical level are indicated. When the use of GIS in agriculture is discussed here, this must be taken in the context of the many agricultural systems, the different political and social structures, the varying surrounding conditions, and the extremely variable farm sizes in different parts of the world. This is illustrated by the example of farm sizes in two European countries: the average farm size in Switzerland is around 14 ha, mainly in smallholdings. In Germany, the average size is around 45 ha with an extreme north–south and east–west differentiation. For instance, in Mecklenburg-Vorpommern, the average farm site is around 250 ha, and some agricultural concerns manage farms larger than 1000 ha.

From this, we can say in advance that most of the statements in this section are related to parts of the *developed* world, in which high-technology agriculture has been adopted and where the use of GIS has become a general matter of course.

The spatial dimension of agriculture is easily identified: cropland, pastures, plantations, and grazing land define the rural landscape in large parts of the world. The overall effects of local and regional environmental conditions on agriculture are also superficially easy to identify, with patterns of agricultural land use dependent on soil type, climatic conditions, availability of water, etc., and thus differing across different regions of the globe. Potential uses of GI technologies in agriculture are also readily apparent. On the one hand, the basic information related to agricultural parcels, such as usage type, yield, and records of operations, such as plowing and fertilization, have an inherent spatial (and usually also temporal) reference. On the other hand, analysis of site-specific conditions and the reaction of crops and animals to these may assist the farmer in more efficient agriculture. At a national and regional level, this is already practiced, e.g., in the production of recommended varieties and fertilization recommendations, but it may also be used at a part-field level to optimize use of resources within individual crop stands.

Currently, there is a vast range of products for agriculture: machinery producers (Full-Liner, Komatsu) and company consortiums offer complete agricultural software solutions, which also include a GIS module, e.g., AGRONET, AgroOffice, JD-Office, Helm, etc. Many small (GIS) vendors concentrate on regional markets with proprietary solutions and special services for farmers, e.g., as information technology (IT) vendors and contractors or as value-added resellers of geographic base data. There is also an increasing trend towards expert systems and specialized web-based applications with GIS functionality (ISIP, Yara Sensor Office, Pro-Plant expert, etc.). However, most of the development of agricultural software solutions is driven by vendors and authorities and has let to non or only partly standardized products and services.

24.1 Spatial Data in Agriculture

Spatial data related to agriculture may be collected on the farm or provided by external agencies, and farmers must often provide spatially referenced data to a range of third parties, in administration, in agricultural support services, and in the value-added chain. Three dimensions of agriculture are, therefore, often postulated (Fig. 24.1), representing different chains through which spatially referenced agricultural data may be exchanged.

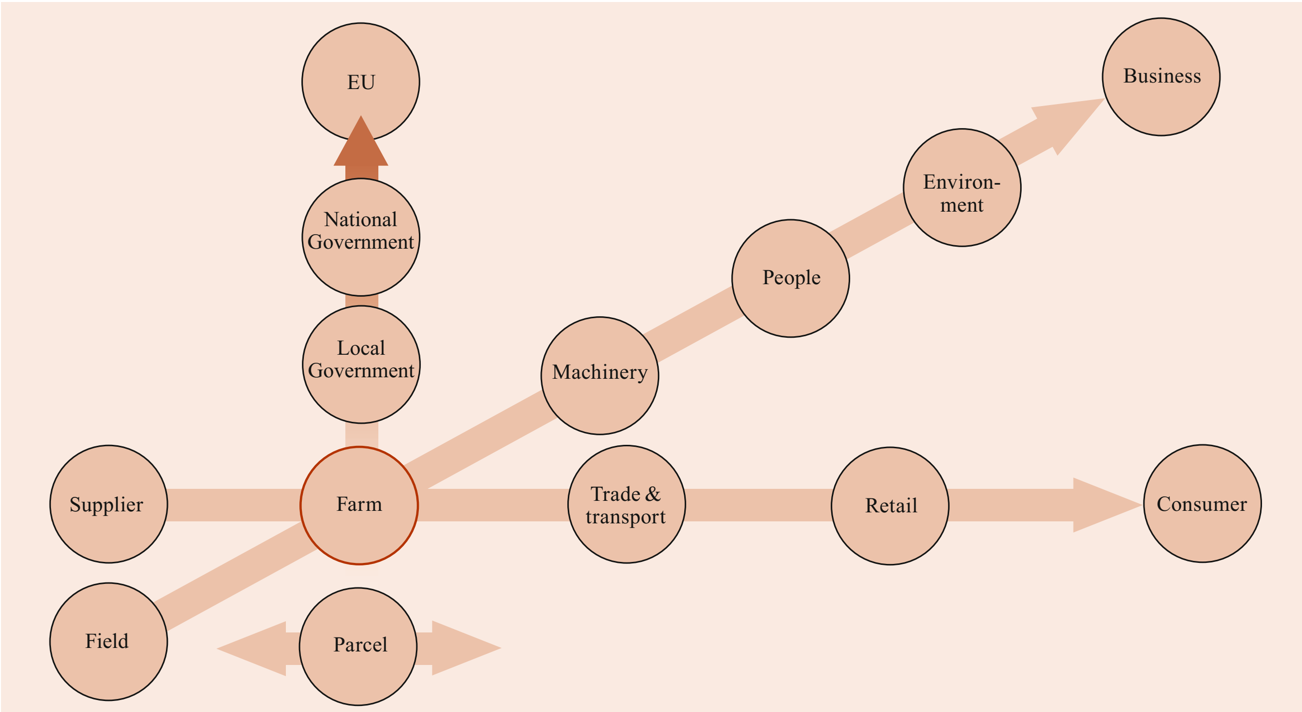


Fig. 24.1 The three dimensions of agricultural data. (After [3])

The use of geographic information in agriculture will be considered here in three categories:

1. GI as a reference system for management of data relating to farms and fields
2. GI as a supporting tool for performing field operations (e.g., driver guidance and autosteering)
3. GI as a source of information for making and implementing agronomic decisions (site-specific or information-driven farming)

Both the second and third of these categories may be grouped together under the umbrella term *precision agriculture* (Sect. 24.3). However, many of these aspects rely on a common data basis, and so we will first consider some of the common data sources for geodata in agriculture.

24.1.1 Data Sources

The data sources, information processing, and decision processes in precision agriculture have been the focus of recent and ongoing research [4, 5]. One key factor in the choice of data sources is the availability of data; e.g., in some countries, such as Germany, nationwide datasets of soil type are available and form a basis for identifying heterogeneity. We will, therefore, concentrate on the available data, based initially on that which is commonly available in Germany. Comparable data are available in most other developed countries.

Basic Geodata from Administrative Surveys

In Germany, the survey and cadastre administrations produce so-called geo base data for business and government. According to their legal commission, these are produced nationwide, currently provided via standardized OGC web services, and as far as possible consistent across the total area of the Federal Republic. Their relevance for agricultural applications is shown in Fig. 24.2. For the management of leaseholds and other farmed land, property rights from the cadastral system are very relevant and are also required for the EU Integrated Administration and Control System (IACS (Sect. 24.2)). Currently, an integrated model of the former real estate map (the parcels, their geometry and usage) and the real estate book (the ownership and rights) into the new Authoritative Real Estate Cadastre Information System (called *Amtliches Liegenschaftskataster-Informationssystem*, ALKIS) is available all over Germany. Topographic maps – as part of the existing Authoritative Topographic and Cartographic Information System (called *Amtliches Topographisch-Kartographisches Informationssystem*, ATKIS) – are often used only as background information for farm overviews, although, e.g., the road infrastructure may be used for fleet management. Administrative boundaries are mainly of relevance for regional and national agricultural statistics.

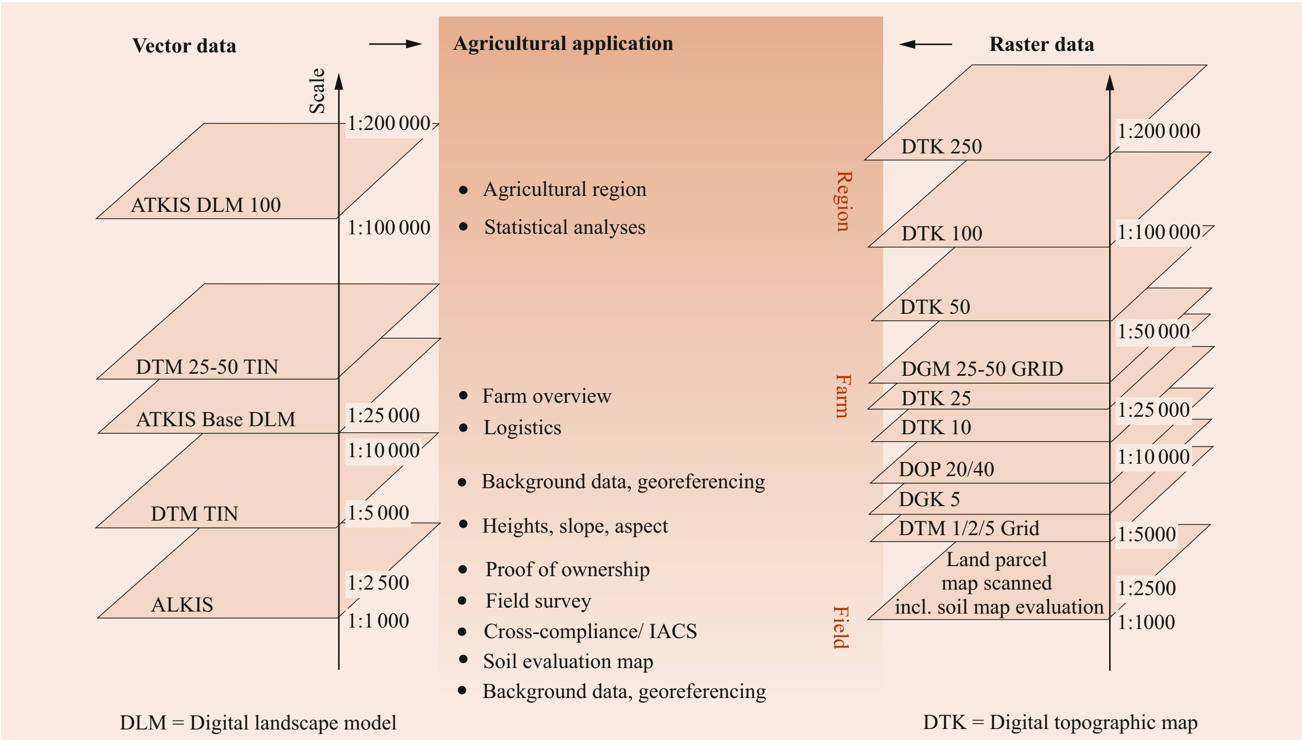


Fig. 24.2 Geo base data and their applications in agriculture. (After [6])

Digital terrain models (DTM) represent the relief of a landscape in digital form. In Germany they are available in different resolutions (from 1 to 50 m grid size) and quality (from dm to some m height accuracy). As the main determining factor for radiation and water balances, the relief controls soil development, the rate of runoff and leaching, material transport, and microclimates, and plays a central role in site-specific differences in agriculture. The DTM may be used for spatial and temporal predictions of the properties of the crop stand and its growth. It may be straightforwardly interpreted, is easily managed by computer, and has relatively high data stability. The relief has influences on the local climate (e.g., differences in sun exposure and cold air flow), lateral material transport, movement of surface water, and abrasion and accumulation of matter. DTM in combination with the official soil evaluation is also of particular relevance, especially in relation to soil management. Soil formation processes and yield differences are strongly correlated with the relief. From the DTM, it is also possible to deduce whether particular machines may be used due to slope angle, and differences in machine performance (dampness, soil compaction) may be interpreted. Currently, there is a trend towards capture of terrain models using digital photogrammetry and/or airborne laser scanning, which may produce data with more details and higher accuracy.

Soil evaluation is based on a Germany-wide method for the description of soils. The original data of the earlier *Reichsbodenschätzung* are held in the field evaluation books (*Feldschätzbücher*) of the fiscal authorities. They are available as an extra layer for the cadastral maps at a scale of 1:2000 to 1:5000 and are often digitized and prepared for GIS use. In some federal states, the information is provided as a free WMS service. These maps show the spatial extent of the soil number (*Bodenzahl*, a measure of estimated agricultural productivity), the main soil type, the soil condition, and the geological background. A further dataset is a description of bore holes, at which soil type, humus content, hydromorphology, and lime content are recorded in two to four layers. Where newer soil mapping is not available, these datasets still represent a good information basis for tillage, sowing, and base fertilization [7].

Specialist Administrative Geodata

Today, nearly all information related to the landscape is represented in map form. Similarly to the governmental cadastral and survey agencies, other state bodies are also tending to deliver information derived from the analogue map materials in digital form.

This is, for instance, the case for the soil, geological, and geomorphological map series, which are gathered and maintained by the state geoscientific agencies nationwide in the so-called soil information systems (*Bodeninformationssystem*, BIS) as a comprehensive and systematic soil inventory. As well as soil science data, the BIS includes data on the geological composition of the Earth's upper crust alongside information on hydrogeology, resilience, engineering geol-

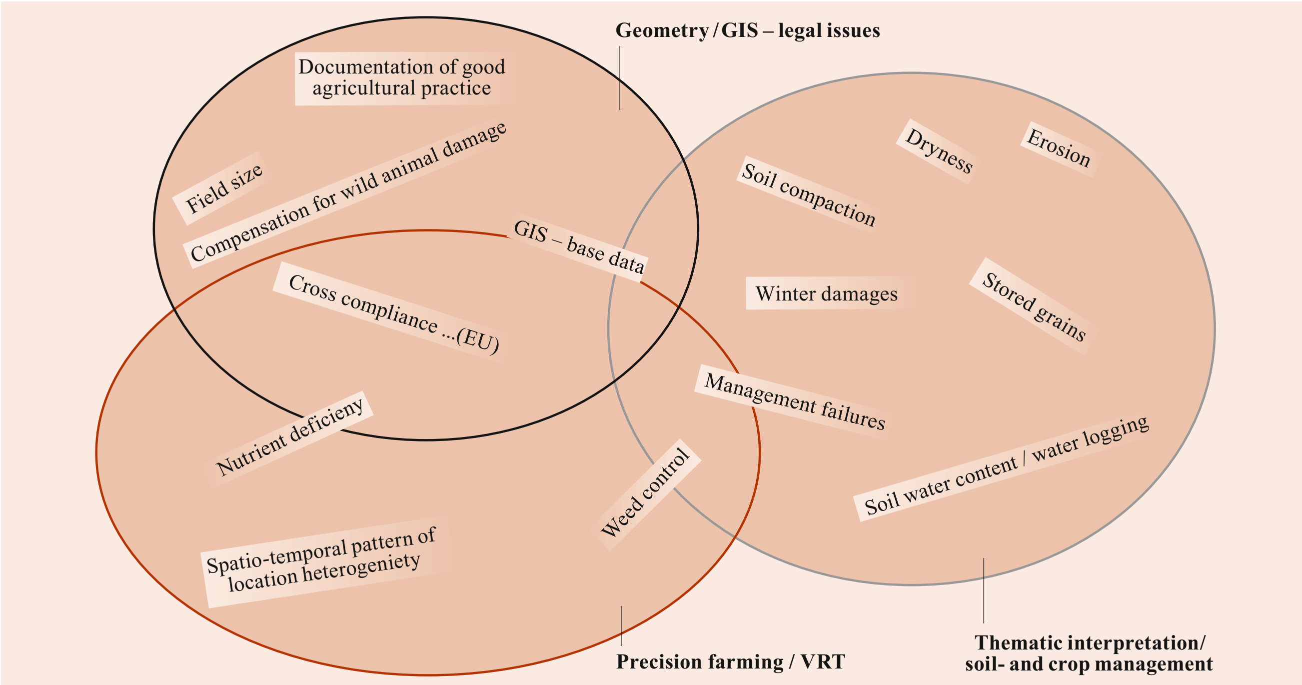


Fig. 24.3 Potential applications of remote sensing in plant production (VRT = variable rate technology)

ogy, and geochemistry. They contain descriptions of the local distribution of soil types and their properties, such as soil structure, humus content, pH value, soil density, parent material, and upper-layer water balance. Soil mapping includes selective sampling through drilling and/or on the basis of test pits, as well as the spatial extents of regions with identical soil properties or, depending on scale, similar properties. Usually the pedosphere is sampled to a depth of 2 m under the surface. Mapping on the basis of probes may, as well as soil type, also show the soil capacity or the erosion risk. The data are represented as probe descriptions, analyses, and thematic maps at various scales. Most of them are made available via standardized web services.

Within the EU-wide Integrated Administration and Control System (IACS (Sect. 24.2)) for the monitoring and implementation of a consistent agricultural policy in the EU member states, the state agricultural agencies now hold contiguous parcel maps of all agricultural land. These are held digitally and are usually available via web services (e.g., in Rhineland-Palatinate at http://www.geoportal.rlp.de/ or in Bavaria at www.geoportal.bayern.de/). IACS is used for calculating subsidy payments to farmers and controls whether the correct crop type was grown and whether the stated extent was correct. The basic unit for the so-called multiple application is the field plot or physical field block.

24.1.2 Remote Sensing

The use of remote-sensing data has a long tradition in agriculture [8]. Both satellite-mounted and aerial sensors are used, as well as sensors mounted on agricultural machinery. While satellite remote sensing is mainly used in the national and international range, e.g., for the prediction of droughts, remote-sensing data is also used at the farm level, e.g., for the determination of parcel sizes for EU subsidy payments and as base data in precision farming. The potential of remote sensing data for farm-level plant production is great, as shown in Fig. 24.3.

As Fig. 24.3 makes clear, remote-sensing information can be used not only for precision farming in the narrow sense but also as a basis for the interpretation, together with advisors and experts, required to answer many questions regarding soil and crop management. However, this information is required in near real time and, where possible, in all weather conditions. Therefore, the use of specialized aerial photography missions or the use of drones or unmanned aerial vehicles (UAVs) is necessary. The use of tractors with special sensors (nitrogen sensor, Crop-Circle, CropSpec, NIRS-sensors) to measure the variability of water and nitrogen content and some other parameters has proved to be helpful for certain applications.

24.1.3 Internal Farm Geoinformation

Parcel Measurements

One fundamental spatial dataset required for use of GI in agriculture is the exact boundary of each field and/or crop stand. Even where digital cadastral or topographic mapping is available to the farmer, it is usually necessary to capture the boundaries, as they may not match land parcel boundaries or the features shown on a topographic map. The accuracy requirements for boundary measurement are defined by law for some purposes, e.g., IACS (Sect. 24.2). Although in some cases the boundaries may be captured using traditional survey methods, two common methods are capture using GNSS, for which the boundary of each field is simply traversed with a high-accuracy GNSS receiver (real time kinematic, RTK), or manual tracing from precise orthophotos, produced by aerial imagery or UAV imagery. In both of these cases, it is possible to capture new boundaries, e.g., where an existing field is planted with two different crops, creating two new crop stand boundaries, with relatively low cost, high precision and good actuality.

High-Precision DTM

As an alternative to the official DTMs from the state survey agencies, it is also possible to produce a *high-precision DTM* using DGNSS (differential GNSS), either using existing farm equipment, such as may be used as the basis for parallel driving systems (in *GNSS Guidance and Autosteer Systems*) or as a service provided by a contractor [7, 9]. Such a survey is completed using two high-quality GNSS receivers with an accuracy of ±2–5 cm. With one receiver at a known point (Chap. 8) or using a reference signal, e.g., from a satellite positioning service (e.g., StarFIRE, SAPOS (German satellite positioning service)), the majority of error sources may be corrected directly in the field (RTK mode) with a height accuracy of 10–15 cm [10]. The height of the terrain is collected during field operations, and this may be combined with other data collection, e.g., EM38, or during normal operations in the drive lanes.

On-Site Collection

Soil samples are, as a rule, collected every 6 years (the minimum legal requirement is one sample per 5 ha every 6 years) in order to determine the soil texture, nutrient content (Ka, P, Mg, etc.), pH, humus content, cation exchange capacity, etc. For the planning of sample locations, there are two approaches: raster sampling and directed sampling. Raster sampling is the standard for precision farming, whereas directed sampling is an interesting alternative where there is already good knowledge of the crop stand. In raster sampling, 20–40 individual samples are combined into a mixed sample, either around a single point, along a diagonal, or in a zigzag pattern. From the chemical analysis, farm soil maps and management zones may be generated to form part of the decision process for agronomic applications.

An alternative approach was presented by Peets et al. (2012) with the use of on-the-go sensors for the collection of soil properties [11]. This approach uses mobile VIS-NIR spectroscopy in the field to measure soil spectra in a high resolution. The measured values are then compared with properties of soil samples to establish a typical soil model, which can then be used to predict soil properties from VIS-NIR spectra data. The number of measurements increases significantly (this is comparable with other GNSS-based field operations) and could serve as a source for precise soil fertilization and liming operations.

During the crop growth season, plant ratings and field measurements are performed, usually supported by GNSS. These may be used, e.g., for the calculation of crop density and the derivation of the total biomass, or for the mapping of weed distribution, which are, in turn, used as the basis for decisions for agricultural operations. Farmers may also perform their own field trials (parcel-based on-farm experiments), which will also be mapped and contribute to the sum of farm-internal data for the management of crop production.

24.1.4 Other Data

Further spatially referenced datasets are also used in agriculture. One example is weather data, which may, on the one hand, be obtained from national meteorological agencies or, on the other hand, be provided by on-farm or regional weather stations serving the agricultural community. The record of past conditions and forecast future conditions play an important role in decision-making, e.g., spraying operations may be forbidden or not possible during rain or high winds, and many operations are not possible while the soil is frozen. In areas with limited water availability, decisions on irrigation and other inputs where the uptake is dependent on sufficient available water, the need for locally generated rainfall records and forecasts are also immediately apparent.

24.2 Integrated Administration and Control System

The Integrated Administration and Control System (IACS) was introduced in 1992 by EU Council Regulation 3508/92. The aim of this system is more effective administration of financial aid payments to farmers and to prevent fraud more easily by allowing crosschecking. This legislation required each member state of the EU to set up a computerized database system comprising an alphanumeric identification system for agricultural parcels based on land registry maps and documents, other cartographic references, or of aerial photographs or satellite pictures, or other equivalent supporting references, or on the basis of more than one of these elements.

Although the requirement was specified to identify agricultural parcels, particularly with respect to support schemes linked to surface area, no requirements were specified with regard to storage and management of spatial data. However, due to the use of GI-based systems in all member states and the increasingly widespread availability of GI systems and data, this regulation was amended in 2000 by EU Council Regulation 1593/2000 to explicitly mandate the use of GIS for the identification of agricultural parcels, specifying that accuracy should be guaranteed equivalent to 1:10 000 topographic mapping.

Furthermore,

> … Member States shall simplify the application process by distributing pre-printed forms based on the areas determined in the previous year and supplying graphical material … indicating the location of those areas.

The original legislation has been modified frequently by many further EU Council Regulations, but the requirement for all EU member states to maintain a GIS database of agricultural parcels, and to make available extracts of this data for farmers, has remained constant. This legislation has, therefore, created one of the largest GIS projects in the EU.

24.2.1 Land Parcel Identification System

This GIS component of IACS [12] is known as the Land Parcel Information System (LPIS). As is normal for EU regulations, each member state may decide how this is implemented, as long as it meets the minimum requirements laid down by the EU. Two key spatial features are defined as: the agricultural parcel (also known as the production unit) and the reference parcel (the production block). The reference parcel is defined in EC Regulation 796/2004 as "… a geographically delimited area retaining a unique identification as registered in the GIS …", whereas the agricultural parcel is defined in EC Regulation 972/2007 as

> … a continuous area of land on which a single crop group is cultivated by a single farmer; however, where a separate declaration of the use of an area within a crop group is required in the context of this Regulation, that specific use shall further limit the agricultural parcel.

However,

> … the identification system for agricultural parcels … shall operate at reference parcel level such as cadastral parcel, or production block which shall ensure unique identification of each reference parcel

(EC Regulation 796/2004).

The reference parcel may, therefore, be semipermanent, which is indicated by the expected features to be used as physical block boundaries [13]:

- Infrastructure (roads, railways, water channels, etc.)
- Farm tracks and other limits between land cover types that are considered mostly permanent (streams, vineyard, orchard/olive grove limits, woodland borders, etc.)
- Limits between parcels of the same cover type that can be considered permanent (fence-lines, hedge-rows, etc.).

However, not all member states use physical blocks as reference parcels; the agricultural parcels themselves, farmer blocks (defined as a piece of land cultivated by one farmer with one or more crops), and cadastral parcels are also used, although as noted by *Léo* and *Lemoine*, "… cadastral maps may be at a large scale and recorded at a high precision, their information is related to land ownership and not to the real agricultural parcels …". Various possible relationships are, therefore, possible between reference parcels and agricultural parcels (Fig. 24.4).

In the 16 federal states within Germany, 13 different solutions are used; these being different combinations of parcel identification systems (lot, block), different spatial reference

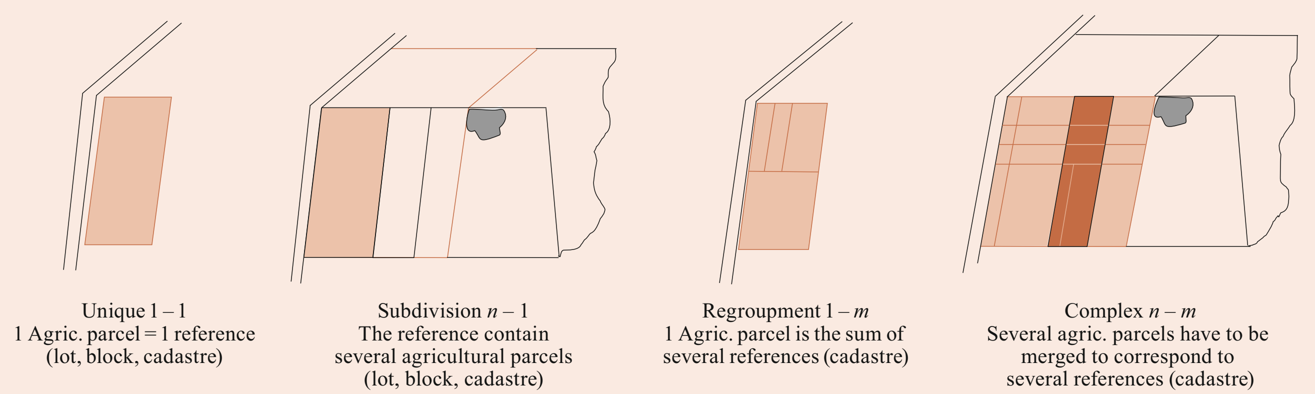


Fig. 24.4 Relationships between agricultural and reference parcels. (After [13])

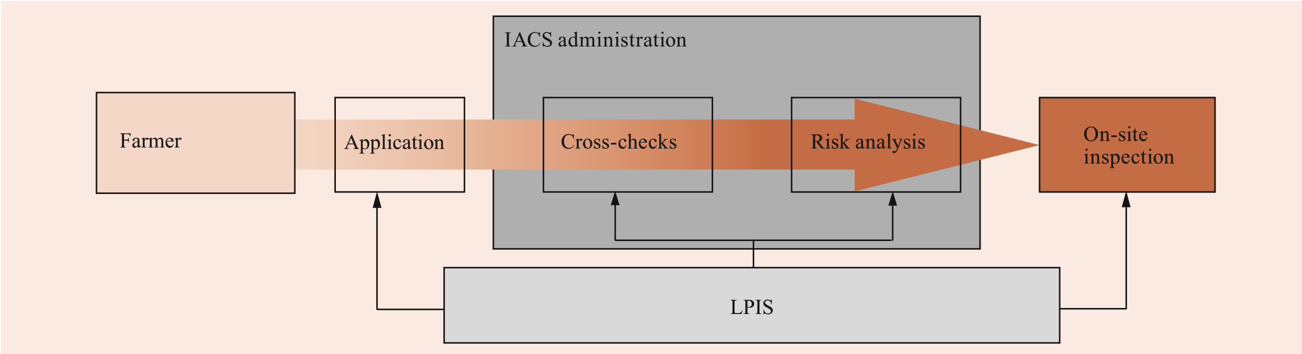


Fig. 24.5 LPIS (land parcel information system) workflow

systems, different subsidy application processes (online/offline), and different on-site controls, which makes the use of digital data for national agricultural services difficult.

The farmer makes the subsidy application at the agricultural agency, for which the LPIS is used for the unique spatial identification of eligible parcels in the EU. In the ideal case, the complete subsidy application, including spatial aspects, may be completed online. For this, some federal states make Internet-capable viewers with basic GIS functionality available to farmers, such that, based on the geo base data (orthophotos/cadastral parcels) from the regional survey agencies, the farmer can see the fields and parcels belonging to the farm and select those to be included in the subsidy application or modify existing parcels where changes are needed (Fig. 24.5).

24.3 Precision Agriculture

24.3.1 Precision as the Basis for Modern Agriculture

Many of the fields around the world are heterogeneous. They have small-scale differences, which result from the influence and effects of variabilities in soil, relief, human, and management factors, as well as being due to the location of the field in relation to other landscape features. These site-specific differences impact on the plants grown, resulting in inhomogeneous crop stands and differences in yield [14]. Using modern technology, such small-scale differences can be taken into account during management and application processes.

Such site-specific, information-driven crop production is usually referred to as *precision agriculture* [14, 15] or *precision farming*, although the exact definition of this term is not universally agreed upon. Similarly, *precision livestock farming* refers to the use of sensors and geoinformation for animal production. In neighboring disciplines terms such as precision horticulture, precision forestry, etc., are used (Fig. 24.6).

Although applications of GI for extensive livestock farming are also under development, particularly the use of GNSS for monitoring animal movements and behavior [16, 17] and virtual fencing [18], the use of GI in the arable farming sector, particularly extensive crop production, e.g., for cereals such as winter wheat, is more widespread and is the primary focus of this section.

Precision farming encapsulates the adaptation of agronomic activities to the variability of the site-specific and crop parameters, which are measured using satellite navigation systems (GNSS, [19]) and combined and calculated in GIS [20]. From this, site-specific agricultural applications and operations may be derived, which are then applied with the help of integrated sensor technologies on agricultural machinery. This can assist the farmer in reducing the quantity of agrochemicals applied, increasing yield reliability and leading towards more sustainable and environmentally friendly agriculture. It has been shown that the greatest potential for these techniques in both economic and ecological terms lies in areas with heterogeneous conditions and large-scale production. It is possible to perform site-specific operations in all field operations, from tillage through sowing to fertilization and plant protection [7]. These spatial technologies are leading to a paradigm change in crop production: whereas previously a crop stand was the smallest unit of crop production, now new spatial technologies allow specific management of subareas within a crop stand.

24.3.2 Spatial Technologies

The enabling technology for precision agriculture is global navigation satellite systems (GNSS), such as the US NAVSTAR-GPS, the Russian GLONASS, the European GALILEO, or the Chinese Beidou/Compass. GNSS receivers are mounted on the tractor, combine harvester, or other field implement. As well as this, precision farming requires a large amount of high-resolution spatial and temporal

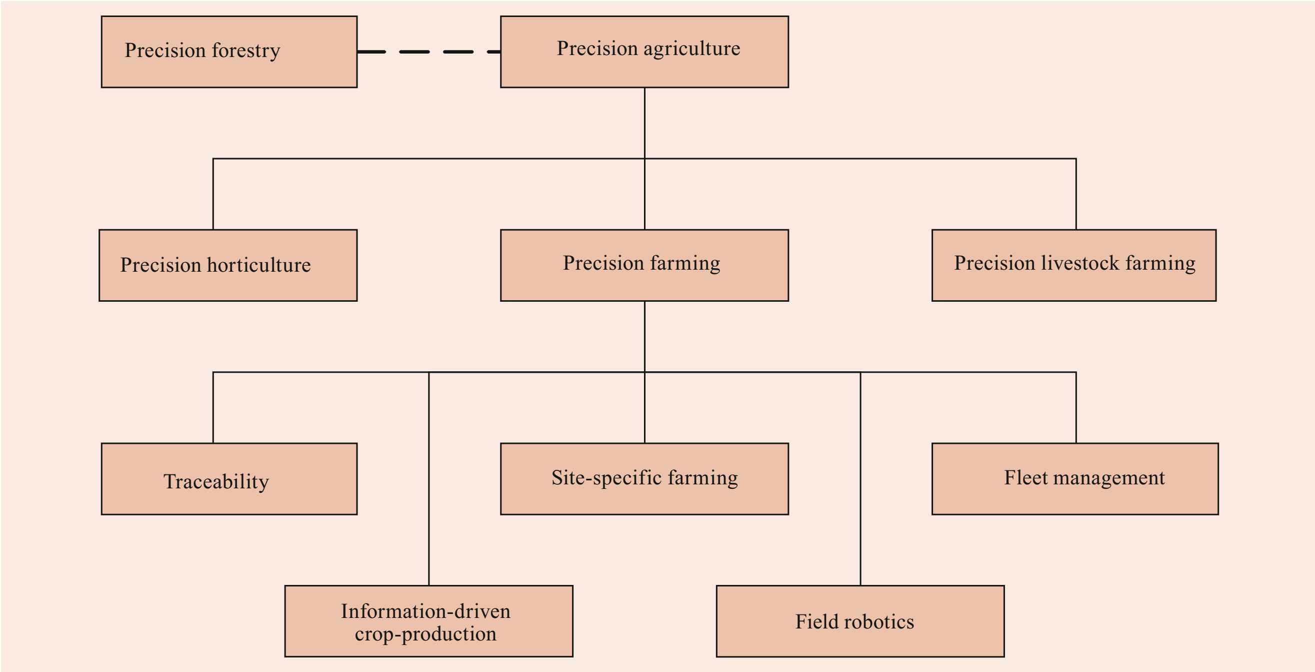


Fig. 24.6 Classification of *precision* technologies in precision agriculture

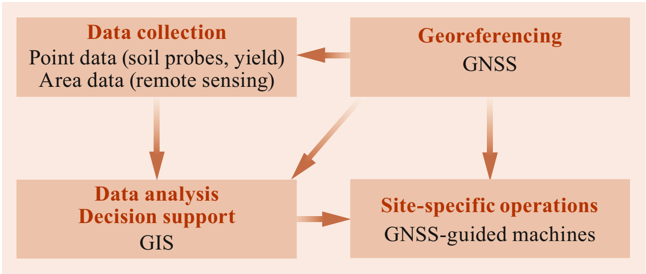


Fig. 24.7 Spatial components of precision farming technologies

data, which must be processed in GIS and applied using the available agricultural machinery (Fig. 24.7).

Usually, DGNSS or RTK-GNSS are used to allow collection of data or application of agricultural chemicals with submeter accuracy. This is combined with agricultural machinery allowing, for applications of fertilizers or plant protection products, exact and variable dosage or onboard sensors continuously measuring yield volume and other parameters. The development of crop growth models and the use of data about the conditions in the field such as soil type, relief, and climate allow optimum inputs to ensure maximum yield or minimum environmental impact to be determined.

The various possible uses for GNSS in agriculture cannot, however, all be realized with a single GNSS solution. In particular regarding the required accuracy (Table 24.1), many different receivers from integrated low-cost devices through to expensive precision receivers must be used.

24.3.3 Precision Farming

As the basis for the further explanation, the precision farming cycle is presented in Fig. 24.8. A variety of agricultural operations are performed during the crop growth period. This starts with tillage, followed by sowing. During the growth of the plants, fertilization and spraying to combat plant diseases and weeds are necessary. As the result of the process, the harvest is collected. Today, many of these operations are performed using machinery, which is equipped with sensors that in combination with GNSS can map the spatially differentiated application of seed, fertilizer, and crop-production products, as well as the resulting harvest yield. In order to plan and perform the individual site-specific operations correctly, further data are required, which must all be analyzed in a GIS or farm management information system (FMIS) [21].

Table 24.1 GNSS accuracy required for agricultural applications

Applications	Required accuracy (m)
Navigation, fleet management, documentation	< 5–10
Yield mapping, soil mapping, observation recording, area measurement for subsidy applications	≈ 1
Parallel driving assistance, terrain modeling	< 0.30
Autosteering	0.02–0.30
Field robotics	0.05
Variable rate technology	< 1

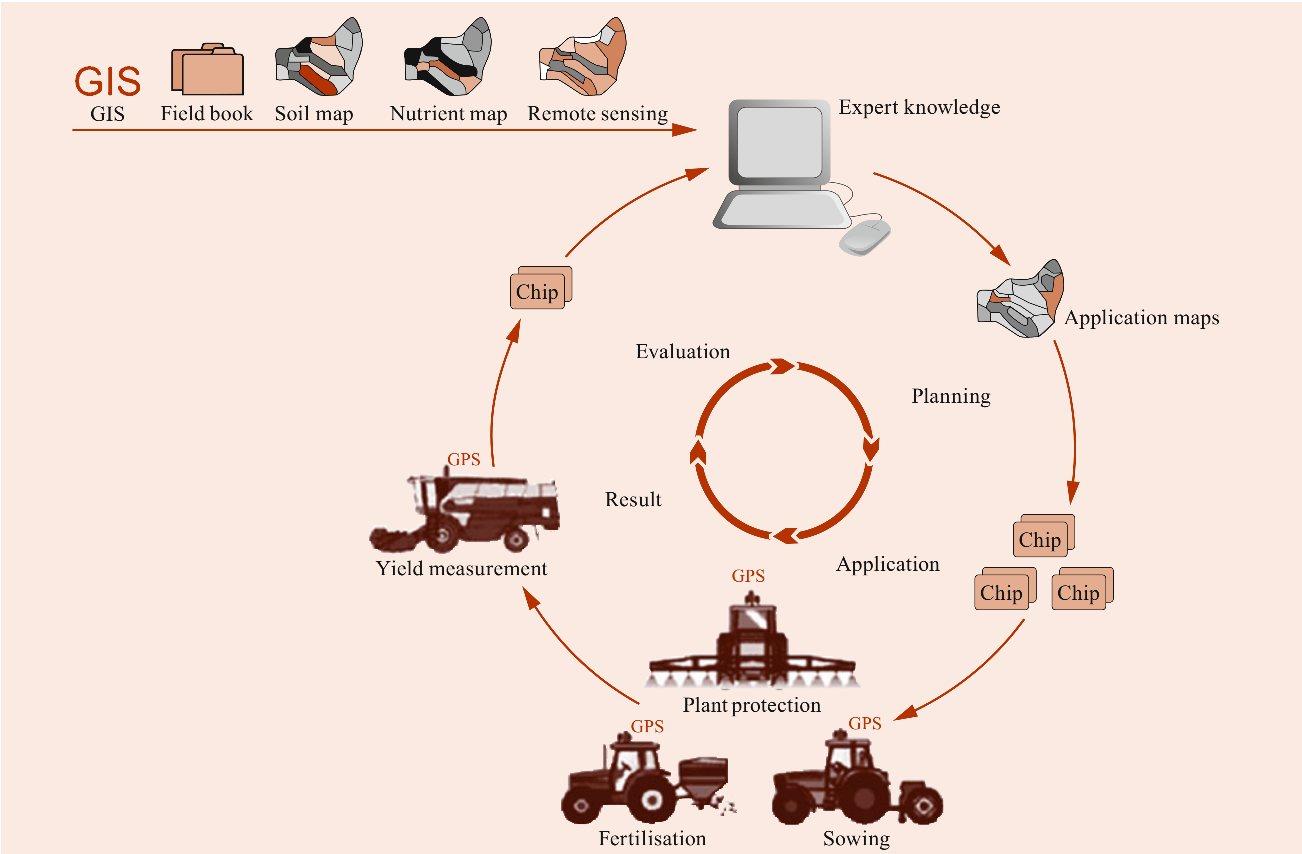


Fig. 24.8 The precision farming cycle. (After [22])

Precision farming can be applied in many different ways (Fig. 24.9). Using the GIS overlay method, a generally static process, the multitude of spatial data are processed using models in order to generate application maps. These are then used to control the field operations, such as sowing, fertilization, and crop protection, and the driver can, usually with only a few presses of a button on the onboard computer, apply more or less or halt the operation. The actual quantities applied are registered and saved on the farm as an as-applied map.

In the online-sensor method – a dynamic method – sensors are mounted on the agricultural machinery to directly sense crop stand characteristics, which are then processed by the onboard computer. The application is then instantly adjusted. This is applied, e.g., for nitrogen fertilization with the Yara N-Sensor, as well as for the application of herbicides and fungicides and for growth regulators.

Both online and overlay methods can be used in a hybrid approach. In this approach, information processing before the operation generates an application map, which is then modified using online sensors and a rule base. Hybrid methods may also be used in N-fertilization and herbicide, pesticide, and growth regulator applications.

GNSS provides one of the foundations for site-specific agricultural management, with which spatially referenced data may be generated and made available in a GIS for analysis. However, a multitude of further information is required in order to implement precision farming technologies across all areas of farm operations. Compared with traditional agriculture, the volume of geoinformation used and the number of tools required to manage this are significantly increased in precision farming [6].

Soil Heterogeneity

There are various methods available to measure soil heterogeneity. For deriving the apparent electrical conductivity (ECa) of soil for agriculture, the contactless close-range remote-sensing method EM38 (Earth conductivity meter) has established itself as a relatively robust and performant method [7]. The apparent electrical conductivity of agricultural soils is strongly related to the clay content but also influenced by the moisture and soil content of the soil suspension. In order to estimate the average clay content of the soil, simultaneous soil core sampling is necessary (in particular to measure soil type and moisture). The probes are used to calibrate the values measured in zones of equal apparent

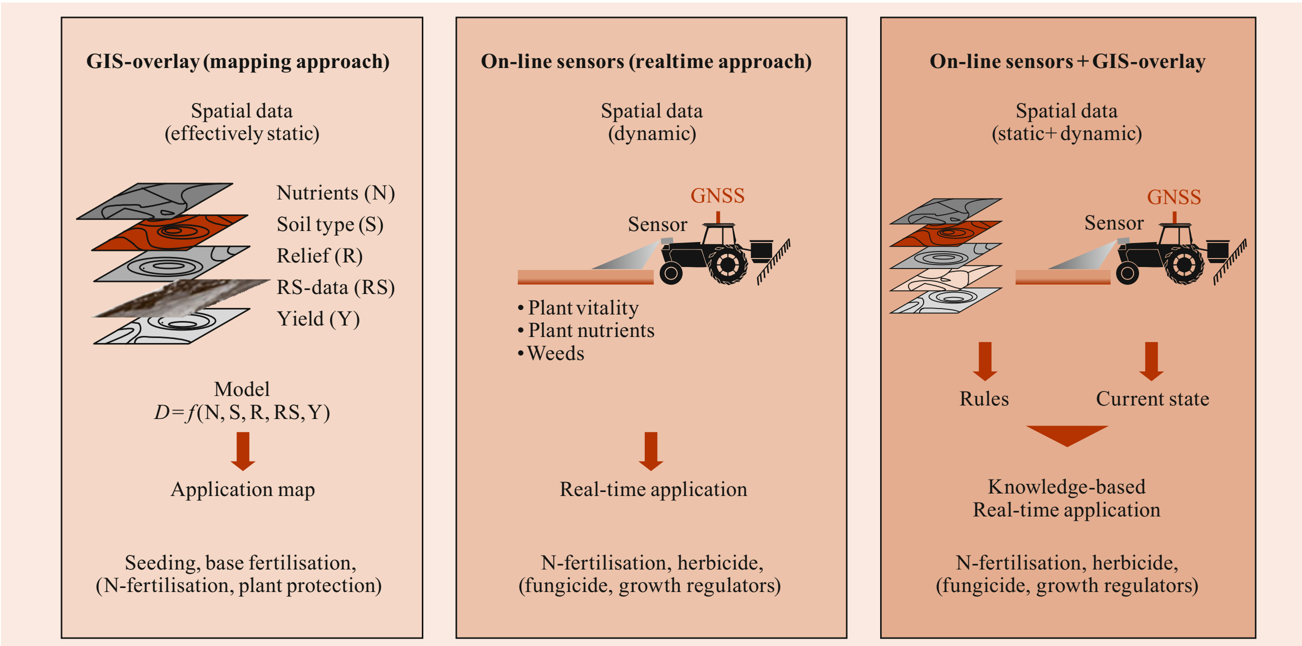


Fig. 24.9 Different precision farming strategies for site-specific applications

electrical conductivity. Measurement of ECa is performed using a sledge towed behind a vehicle. The sensors react very sensitively to metallic objects and underground high-voltage cables.

For site-specific management, usually *raster soil sampling* is used to measure parameters such as soil texture, nutrient content (Ka, P, Mg, etc.), pH, humus content, and cation exchange capacity. The soil samples are analyzed in a laboratory, and the values may be used as inputs to decision-support algorithms. Raster sampling can be used for identification and separation of management zones. This is, however, only sensible in combination with other data (aerial imagery, ECa maps, yield maps, etc.), and knowledge of the field/farm and/or the region are also required. The soil sampling density should be below the desired size of management zones. A higher sampling density is required in areas in which a high variability is expected, and a lower density where low variability is expected. The crop stand is separated into zones, which will be uniformly sampled. Ideally, there are predefined yield or management zones. The shape and size of the zones is determined by the variability and the required operating precision as well as practical and financial considerations. The sampling of the crop stand should concentrate on representative areas within each zone. Transitional regions and anomalies with each zone should also be considered.

Digital farm soil maps represent an important instrument in precision farming for collection, management, and analysis of soil properties, function, and nutrients, with the assistance of GIS. The different soil and nutrient data are the input parameters for the part-field operations performed [7]. These data may be quickly and efficiently collected through use of a *SoilRover* vehicle, collecting and analyzing probes from 1.5–3 m depth, the results of which may be directly entered into a database system and evaluated automatically. In order to minimize the number of probes, further soil sensors, such as ECa measurements using EM38 [23] may be applied, spectrophotometric data (humus quantity of the topsoil) and digital photos collected, and soil resistance (compaction) measured. The sensor data may be prepared in a field laboratory and combined and analyzed with additional existing information (relief, remote sensing, soil estimation, yield) in GIS. For the data values to be generated, differing weightings may be used in a modeling process. Digital farm soil maps are not only product maps but also a complete soil information system for collection and analysis of soil data in precision farming [24].

High-resolution approaches such as the measurement of soil properties with VIS-NIR spectroscopy in the field [11] also require additional support for data collection and management, as presented by *Peets* et al. [23]. Another aspect is the quality of the predicted properties, which can be improved by the utilization of suitable models. A machine-learning-based prediction of precise soil maps to improve the results of VIS-NIR spectroscopy measured with on-the-go sensors was introduced by *Morellos* et al. [25] (Fig. 24.10).

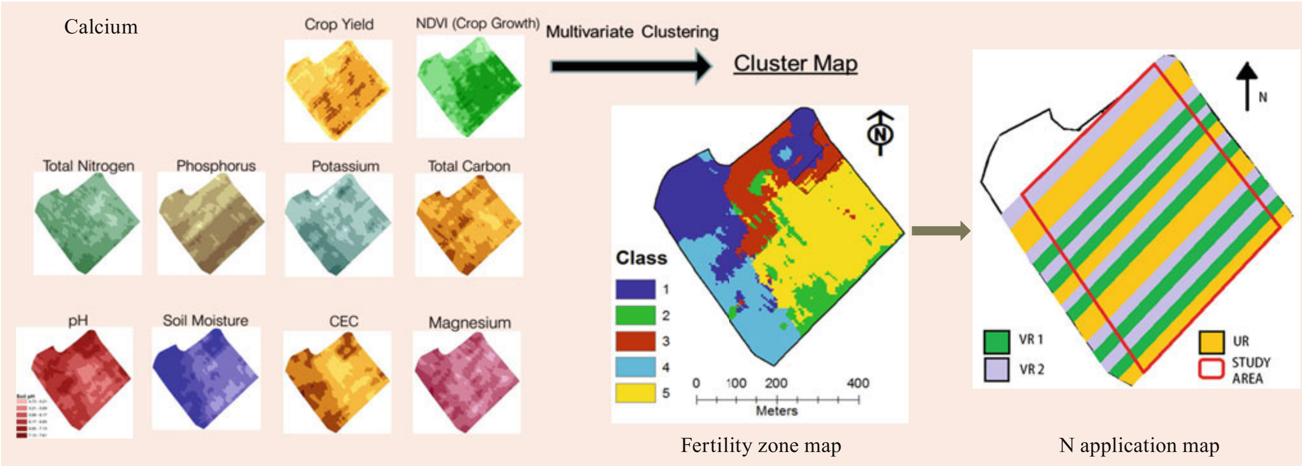


Fig. 24.10 Soil nutrients derived from a VIS-NIR spectroscopy sensor

Yield Mapping

Most farmers who already practice precision agriculture or are planning to introduce it start with yield mapping [26]. Almost every new combine is equipped with GNSS and a yield monitoring system. Consequently, the availability of multiyear yield maps as a basis for delineation of management zones is useful. Accurate yield maps depict the influences of site, climate, and management factors on yield formation for a specific year. Multiyear yield maps may contain valuable information about site-specific yield variability.

In order to produce a yield map that accurately represents yield, some processing is required. Data is collected as a georeferenced time series based on a flow sensor mounted at a point in the harvester mechanism. Typically, the series will also include points measured while the harvester is turning in the so-called *headlands* and thus not harvesting, as well as some unreliable measurements, as well as large random variation due to measurement errors. Furthermore, the measurement points are not distributed evenly across the field but represent an average yield across the strip harvested from each drive line. The speed of the harvester and overlap between the harvest strips will also affect the measured values [27].

In an initial step, the individual point measurements must, therefore, be filtered to remove outliers and adjusted to account for overlap (Chap. 2), [26]. The point values must then be smoothed and interpolated (e.g., using kriging or inverse distance weighting [20, 28]) to produce a yield surface for the field. Once this surface is generated, it must be interpreted in order to make management decisions for subsequent crop cycles. *Blackmore* et al. [29] showed that spatial yield trends do not remain stable through time due to the complex interplay between many factors (e.g., soil, crop type, and prevailing meteorological conditions during the crop cycle), and that fields should instead be managed based on the conditions and variability measured in each given year rather than based on historical yield data. Despite its initial promise, the use of yield maps is, therefore, regarded as problematic in practice as a basis for decision-making [26], and soil maps are often preferred [30].

Terrain Modeling

Relief influences the process of soil formation, water balance, microclimates, and thus the yield capacity of the soil to a large degree. The consideration of relief in site-specific management through use of digital terrain models is, therefore, an important method, particularly as it is based on easily gathered base data with wide application potential for precision farming.

Nowadays, national and regional DTMs are available. In general, a raster size of 5 m with a height accuracy of <0.1 m is required for most agricultural purposes; consistent and high-quality collection and interpolation methods are fundamental [31]. In Germany, the DTM 1 (1 m grid size) and DTM 2 (2 m grid size), being derived from airborne laser scanning (light detection and ranging, lidar; Chap. 9) and producing a high point density and very good height accuracy of 10–15 cm, fulfill these requirements.

As a method for collecting a DTM at the crop-stand level, parallel drive systems (in *GNSS Guidance and Autosteer Systems*) may be used, having an integrated RTK GNSS receiver that may be able to deliver height measurements with an accuracy of 5–10 cm. However, due to the suboptimal geometry of the point distribution (measurement points at small intervals along driving lanes and with 18–24 m between lanes), measurements must be interpolated with complex algorithms such as Kriging in order to reduce errors in interpretation between driving lanes [10].

Therefore, both GNSS and laser scanning provide a basis for the delivery of high-quality DTMs at scales relevant for agriculture. Analysis of the terrain model is performed using special software and delivers basic information such as slope, exposition, curvature, and inflow area, as well as special relief parameters. Following interpolation of the gathered data into a continuous digital terrain model, various index algorithms may be applied in GIS [31, 32]. The topographic wetness index, $\mathrm{TWI} = \ln(A_s/\tan\alpha)$ (A_s = specific upstream area of a point, i.e., the area from which it is calculated that water will flow through that point, $\tan\alpha$ = local slope) describes how strongly an area is affected by inflow and outflow of water and enables identification of moist and dry areas based on the combination of specific upstream area A_s and slope α. This takes into account that water runs off faster in more steeply sloping areas. The result is the potential pattern of soil moisture after precipitation and the run-off lines along which the movement of water and material will occur. The TWI usually has a strong correlation with soil moisture values obtained by remote sensing or EM38 measurements [31].

A variation on the TWI is the stream power index, which describes the potential abrasive power of the water for every cell in the DTM and, therefore, predicts the patterns of erosion. Such secondary relief parameters are usually based on a combination of basic parameters and empirical or process-oriented formulae. Many equations deliver only potential patterns representing the contribution of the relief in determining the observed process and, therefore, relative values, which may be combined with other spatial data in order to explain patterns of crop-stand heterogeneity [31].

For the calculation of the general soil erosion formula after *Wischmeier* and *Smith*, the length slope factor $(A_s/22.13)^{0.6}(\sin\beta/0.0896)^{1.3}$ is important.

Using a GIS, the relief parameters may be compared with soil, climatic, or yield parameters and further investigated [31]. Relief-based increases or reductions may be considered during the creation of nearly all application maps.

Remote Sensing

A fundamental requirement for successful precision farming is that the heterogeneity of the soil and the crop status must be measured for consideration during the decision-making process. Remote sensing is ideal for this (Chap. 10), as with a bird's-eye view, farmers can get continuous detailed information about their crop stands and then apply their knowledge to react to this. Remote sensing is an indirect method, which, depending on the timing of the data capture, delivers information on differences within a crop stand. These differences may be caused by many factors, such as heterogeneity of the soil, the crop, the nutrient supply, the exposition, the management, etc.

The many possible causes for spectral differences significantly hinder the application of remote sensing, since plants may react to different stress factors (e.g., nutrient deficiency, water deficiency, plant diseases) in – from a spectral point of view – very similar ways.

Additionally, the exact time of data capture plays a central role. On the one hand, the reflective properties of the vegetation and soil, as well as the architecture of the crop stand, change continuously. On the other hand, changes in reflection may be reliably correlated with the desired biophysical characteristics, e.g., leaf area index (LAI), biomass, chlorophyll content, etc. Many common indices, e.g., normalized difference vegetation index (NDVI) reach saturation with increasing leaf area and, therefore, cannot be used for further differentiation.

Remote-sensing data are snapshots of plant development with a limited half-life value. Therefore, it is of utmost importance to obtain the information at the most appropriate development stage of the crop.

Remote sensing for precision farming is performed using many different methods from various sensor platforms. The wide range of sensor platforms, which range from satellites through to tractor-based systems, all have various weaknesses and strengths regarding availability, ground resolution, etc. In the overview presented in Table 24.2, the different sensor platforms and their main uses are described and compared.

From the farmer's point of view, particularly *terrestrial tractor-based sensors* are interesting, as these can deliver weather-independent on-the-go spectral information for agricultural operations and, therefore, generate a direct benefit for precision farming.

In order to measure plant vitality, in particular with a view towards nitrogen delivery, many tractor-based systems have been developed in recent years, which are based on differing optical or mechanical techniques:

1. The *Yara N-Sensor* measures the reflective properties of plants under different viewing angles in the visible and near-infrared (NIR) spectrum in the vicinity of the tractor. Based on a large experimental data base and in-field calibration, N-fertilization recommendations are instantaneously available.
2. The crop circle and other scanners, such as ISARIA or CropScan, provide an active light source, which is kept close to the crop stand. The registered reflection at a selected wavelength is independent of the weather conditions. Calibration and N-fertilizations ratings must be determined by the farmer.

The use of *unmanned aerial systems (UAS)* has recently become a focal point for research [32], because these systems are able to gather up-to-date remote-sensing data for small areas, independently from the weather conditions. Depending on the camera used and the flight height over the terrain, UAS images can observe crops at the so-called leaf level or the canopy level. Images taken at only a few meters over the ground at the leaf level provide answers to weed detection

Table 24.2 Comparison of various remote-sensing sensor platforms for precision farming applications

	Terrestrial	UAS/Airborne	Satellite
Platform (sensor)	Tractor (N sensor/green seeker/Crop-Spec/VIS-NIR/laser fluorescence)	Rotary and fixed wing UAS	Sentinel 2a/b/Landsat 8/Planet labs/Landmapper
Application area (strengths and main use of sensors)	Operational support (N fertilization, fungicide etc.)	Operational support (Crop height, N fertilization, fungicide, etc.)	Operational support/base data (culture type, yield potentials, time series analysis, hail/damage mapping)
Information requirement	Real time	Near real time/short term	Short term/strategic
Regionality	Crop stand	Single crops/crop stand	Crop stand/regional
Weather independence	+++	+++	+/++
Turnaround time	Instant to 48 h	1 to 48 h	12 h to 5 days
Spectral resolution	Selected narrow spectral range (10–20 nm), active sensors	RGB, multispectral, (hyperspectral)	Broad spectral bands (50–200 nm)
Atmospheric effects/correction requirements	Low/low	Low/medium	High/medium
Geocoding workload	Low	High	Low
Price (crop stand)	+	+	+
Price (region)	+++	+++	+
Spectral calibration	+++	++	+++
Analysis (derivation of end-products)	Automatic/farmer	Farmer/expert	Automatic/expert

and crop diseases, etc., whereas common canopy level applications are devoted to fertilization and water stress.

UAS generally fly at low altitudes, typically acquiring many images in a systematic manner. With recent developments in computer vision and digital photogrammetry images are mosaicked automatically together to produce a continuous orthoimage of the field or farm of interest. Fixed-wing UAS can cover larger areas because they generally have a longer flight time than multicopters. Common cameras used on UAVs range from inexpensive digital cameras that provide RGB-information to expensive multispectral cameras that provide narrowband reflectance in the blue, green, red, red edge, and NIR regions of the spectrum. Despite promising options for the determination of crop diseases, N-status, etc., the use of miniaturized hyperspectral cameras is currently restricted to research projects due to the complexity of the data analysis and sensor price [33].

Promising results have been obtained using UAV-based remote sensing for estimating crop LAI, biomass, nitrogen status, water stress, weed infestation, yield, and grain protein content. Photogrammetric image processing of UAS data delivers highly accurate 3D point clouds and digital surface models with centimeter accuracy. By subtracting digital surface models of the crops with a reference terrain model of the bare soil, crop height can be determined with high accuracy [34]. Furthermore additional and specific crop parameters can be derived using image analysis tools. As an example, the orthoimages of a mature cauliflower head can be separated from the rest of the crop by its white color. With a raster to vector conversion, the diameter of the approximately round head is available with an accuracy of a few millimeters. Consumers prefer curved heads. This parameter is derived from the ratio of the crop height from the center of the head versus the edge of the cauliflower head (Fig. 24.11).

Spaceborne sensors cover large areas and, with fully automated processing procedures, allow the capture of information at little to reasonable cost. In order to obtain quantitative information about the Earth's surface and to make optical remote-sensing data capable of being spatially and temporally compared, it is necessary to correct for the influences of the Earth's atmosphere. Following atmospheric correction, bio and geophysical parameters such as leaf area index, proportion of photosynthetically active radiation, etc., may be derived and modeled. In recent years, two hardware developments boosted the market for spaceborne remote sensing for agricultural purposes.

1. The European Union launched Sentinel 2a/2b satellites in 2015 and 2017. These satellites provide free multispectral data at 10–20 m resolutions with a repetition rate of 5 days.
2. Private companies such as Planet Labs launched a huge swarm of approximately 150 small earth observation satellites allowing for a daily coverage of the whole Earth. Astro Digital will follow with another 30 satellites in the next years.

With the high number of satellites available, a daily coverage with high-resolution data with GSD of less than 4 m is possible. In turn, a more or less continuous monitoring of crop growth at the field level is possible. Site-specific forecasts of the needs of the crops in terms of nutrients, crop protection, and water is key for further savings and higher yields.

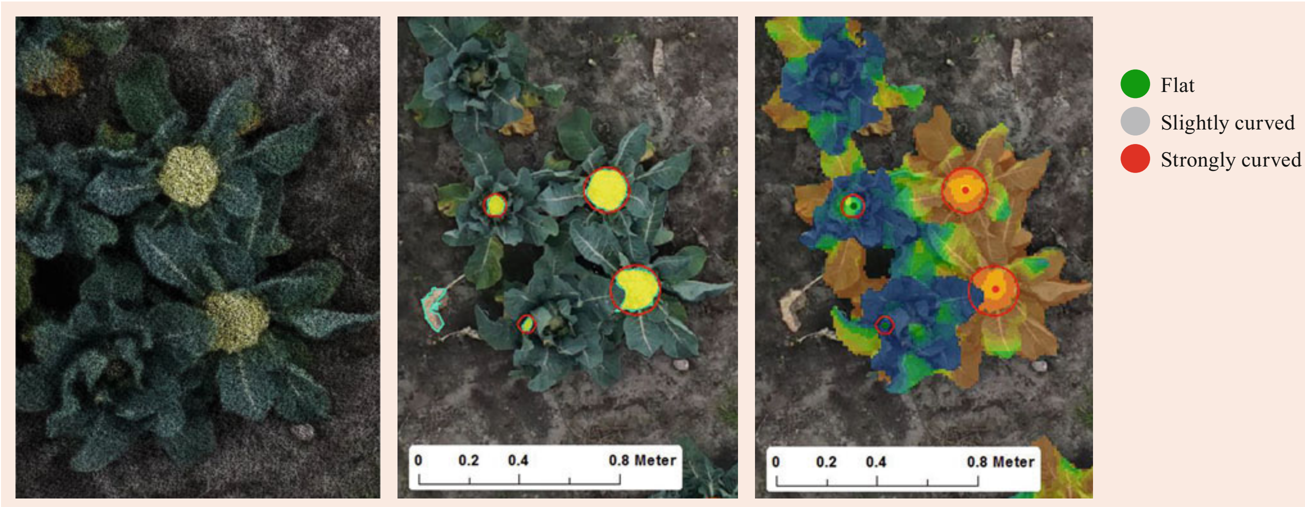


Fig. 24.11 3D-point cloud of cauliflower at leaf level, derived from UAS images taken at an altitude of 20 m above ground and agromarketing of relevant parameters (diameter, curvature)

For better forecasts, a combination of current remote-sensing data with crop growth models is object of many worldwide research activities.

Management Zones

Because influences and interdependencies of factors determining site-specific yield are complex and not always understood, straightforward approaches for the delineation of zones with similar yield potential, and which can, therefore, be similarly managed, are a tool for simple and effective precision agriculture. Up to now many approaches for the delineation of management zones (MZ) have been discussed. *Whelan* and *McBratney* [35] categorize approaches into five groups:

1. Hand-drawn polygons based on yield maps or imagery
2. Classification of remotely sensed data
3. Identification of yield stability patterns across seasons at fixed monitoring points
4. Fuzzy multivariate cluster analysis using seasonal yield maps
5. Morphological filters or buffering.

Once management zones have been delineated, they may form the basis of decision-making for one or multiple field operations, depending on the data sources used and the interpretation of the zones (Fig. 24.12). The management zones here have a size of 30 m by 30 m according to the standard working width used in the farm and are aligned to the standard driving direction. The example shows five different classes ranging from poor soil fertility (red, orange), medium soil fertility (yellow, green), to high soil fertility (blue).

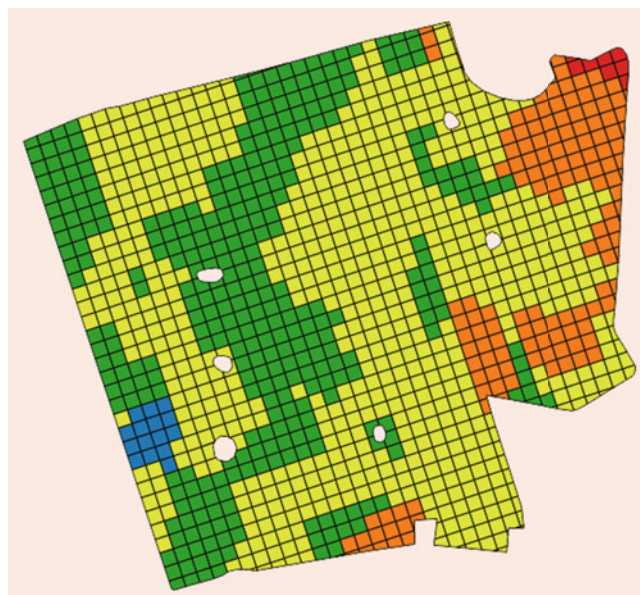


Fig. 24.12 Management zones

GNSS Guidance and Autosteer Systems

GNSS guidance and autosteer systems assist drivers to keep to the desired driving lines during field operations (Fig. 24.13). This leads to a reduction in the overlap of the area being worked, thus reducing the amount of agrichemicals applied and the area affected by compaction from the vehicle's tires. Further economic benefits are gained from a higher average driving speed, reduced fuel consumption, less driver fatigue, and reduced driver and equipment working hours [36]. The results show that there are significant reductions in operational costs varying from 9 to

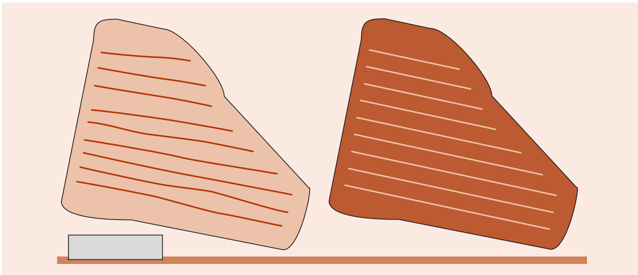


Fig. 24.13 Typical driving lines without and with a parallel driving system (schematic)

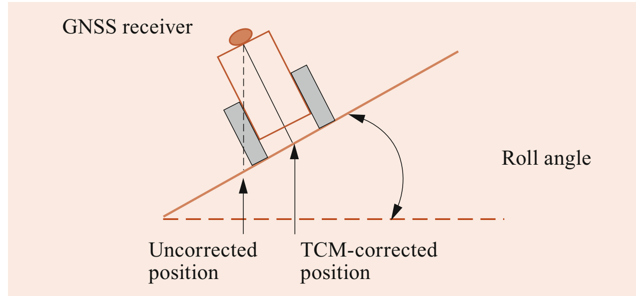


Fig. 24.14 Functionality of a terrain compensation module (TCM)

20%, depending on the specific machinery and field configurations. Such results show the considerable potential of advanced route planning designs and further optimization measures [36].

Usually, driving lines will be parallel, but some systems also allow contour driving or other functions.

Three different levels of assistance exist: GNSS guidance, steering assistance, and autosteer. GNSS guidance gives the driver a visual indication of whether the correct course is being held using so-called lightbars. Steering assistance physically assists the driver in holding the line once the correct path has been entered, usually through hydraulic connection to the steering wheel or a motor with a friction roller on the steering wheel itself. Autosteer is fully integrated into the vehicle and almost completely automates steering. Table 24.3 gives an overview of the different systems and their application areas.

In order to improve the accuracy of parallel driving in undulating terrain, and to compensate for the shift in horizontal position of the roof-mounted GNSS antenna due to vehicle roll on slopes, a so-called terrain compensation module (TCM) may be used. This projects the position measured by the receiver to the true ground position of the center of the tractor using a gyroscope system (Fig. 24.14). The use of a TCM also allows a direct comparison of the positions and heights measured using onboard systems with those in external DTMs.

24.3.4 Information-Driven Plant Production

In order to produce a complete documentation of the production and quality of agricultural products, including all operations performed and all materials used along the complete value-added chain, agriculture is turning towards *information-driven plant production*. The information is not only used for operational planning in precision farming but also offers the opportunity to deliver appropriate information for quality management and controlling along the entire agricultural process chain, e.g., in order to derive process and product indicators. The information may also be used for certification and product liability towards processors and traders. The resulting financial benefits of information-driven plant production result from greater efficiency due to:

- A complete quality-oriented production system
- More transparency in machine use on large farms, machinery syndicates and contractors, and automated contracting
- New performance and person-related billing procedures
- Automated gathering of all crop-stand-relevant management data in a single file
- Complete documentation, e.g., in order to fulfill the requirements of EU Regulation 178/2002 related to continuous documentation on the production and quality of agricultural goods (traceability), including all required operations and applied products

Table 24.3 Overview of the accuracy and applications of different driver assistance systems. (Source [37])

	Driver guidance	Steering assistance	Autosteer
Accuracy	ca. 30 cm with DGNSS	30–5 cm with Omnistar HP, or StarFIRE SF1	Up to ca. 2 cm with RTK
Applications	Lime organic fertilizers tillage	As for guidance, also harvest and potentially sowing	As for steering assistance, plus sowing
Driver relief	Low	High	High
Steering	Manual	Automatic	Automatic
Price (net)	From 1800 to ca. 7500 €	From 9000 to ca. 19 000 €	From 8500 € (DGNSS) up to ca. 40 000 € (RTK)

- Better internal auditing through (partial) crop-stand-specific balances
- Last but not least, simplification for the farmer of subsidy applications and the many other communications with the outside world.

Additionally, ecological benefits are achieved, such as effective integration and documentation of environmental protection goals from water and contract nature protection schemes where farmers are paid for conservation of nature, e.g., part-field-specific documentation of the appropriate use of fertilization and crop protection agents (e.g., optimization of nitrogen efficiency and minimization of applied nitrogen), documentation of additional expenditure for water protection for cross-compliance, and organic fertilization (slurry application plans). Farmers increase their chances of certification or of selling their products to particular markets by meeting particular quality requirements.

In particular, two aspects will be presented here, namely how data are captured on agricultural machinery and transferred for further use on-farm, and how on-farm processing of data may change due to the influence of the spatial data infrastructures that are currently being constructed.

Information Gathering on Agricultural Machinery

Agricultural machinery is one of the most important information sources for collection of in-field data (Fig. 24.15). Through measuring the site-specific yield during harvesting, the effects of agronomic decisions may be analyzed, and through recording the exact quantities of fertilizer and plant production products applied, the requirements of environmental protection and traceability may be met. Furthermore, the performance of agricultural machinery may also be assessed through recording fuel consumption, motor speed, etc., and this information may even be transmitted and analyzed in real time with telematics systems to prevent expensive and time-consuming equipment failures.

For useful analysis of much of this data, it is essential that it is georeferenced, and it is desirable that all data be logged and simultaneously georeferenced by a central system. If problems of proprietary lock-in where all equipment must be supplied from a single vendor are to be avoided then this requires that all devices attached to the vehicle can be connected and communicate in a standardized manner, requiring both hardware and communications/software standards. Examples for relevant standards are NMEA2000 for GNSS technology and the ISO 11783 series (ISOBUS) for agricultural machinery communication.

ISOBUS is based on the principle of a task controller, which forms the hub of the onboard system. This task controller communicates with all onboard devices, including positioning receivers, over a bus system. Within the onboard system, each device (e.g., a sprayer head) is located at a known three-dimensional (3-D) position in a local platform-centric coordinate system. Using the measured real-world position and heading information of the platform (e.g., the tractor), the real-world position of each individual device can be determined. A preprepared application map can thus be implemented and/or spatially referenced data recorded. The transfer of data between the task controller and the farm software is performed using an extensible markup language (XML) file format. Within this format, spatially referenced data may be encoded using either a simple vector or grid-based model (Fig. 24.16). It has to be stated that many current ISOBUS implementations on the machinery only support the grid-based model of application maps. This model is not suitable for field operations with high-precision requirements, e.g., the application of plant protection products in the neighborhood of sensitive areas. Therefore, it would be desirable to have more implementations of the ISOXML vector-based model, which would also improve the interaction with vector-based services of spatial data infrastructures (Sect. 24.3.4, *Spatial Data Infrastructure for Agriculture*).

Information-gathering on agricultural machinery as a whole is an on-the-go process without much interaction of the operator. The new challenge is to make this huge amount of information count for farmers and consumers.

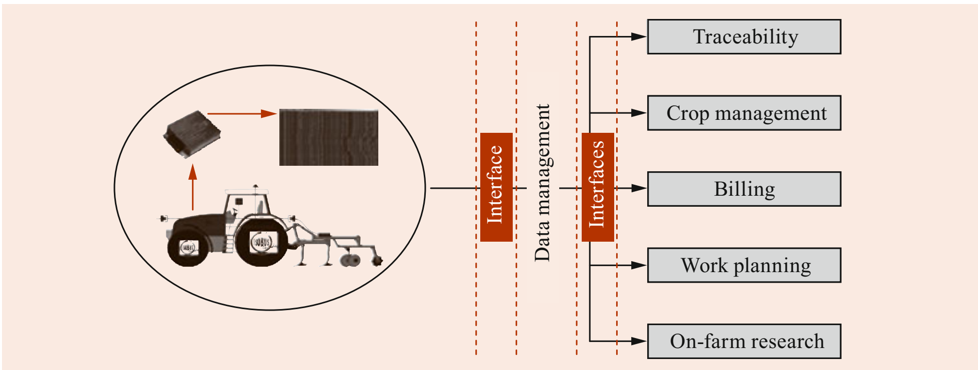


Fig. 24.15 Logging and documentation of all information georeferenced during the application, followed by transfer to agricultural software to support various tasks

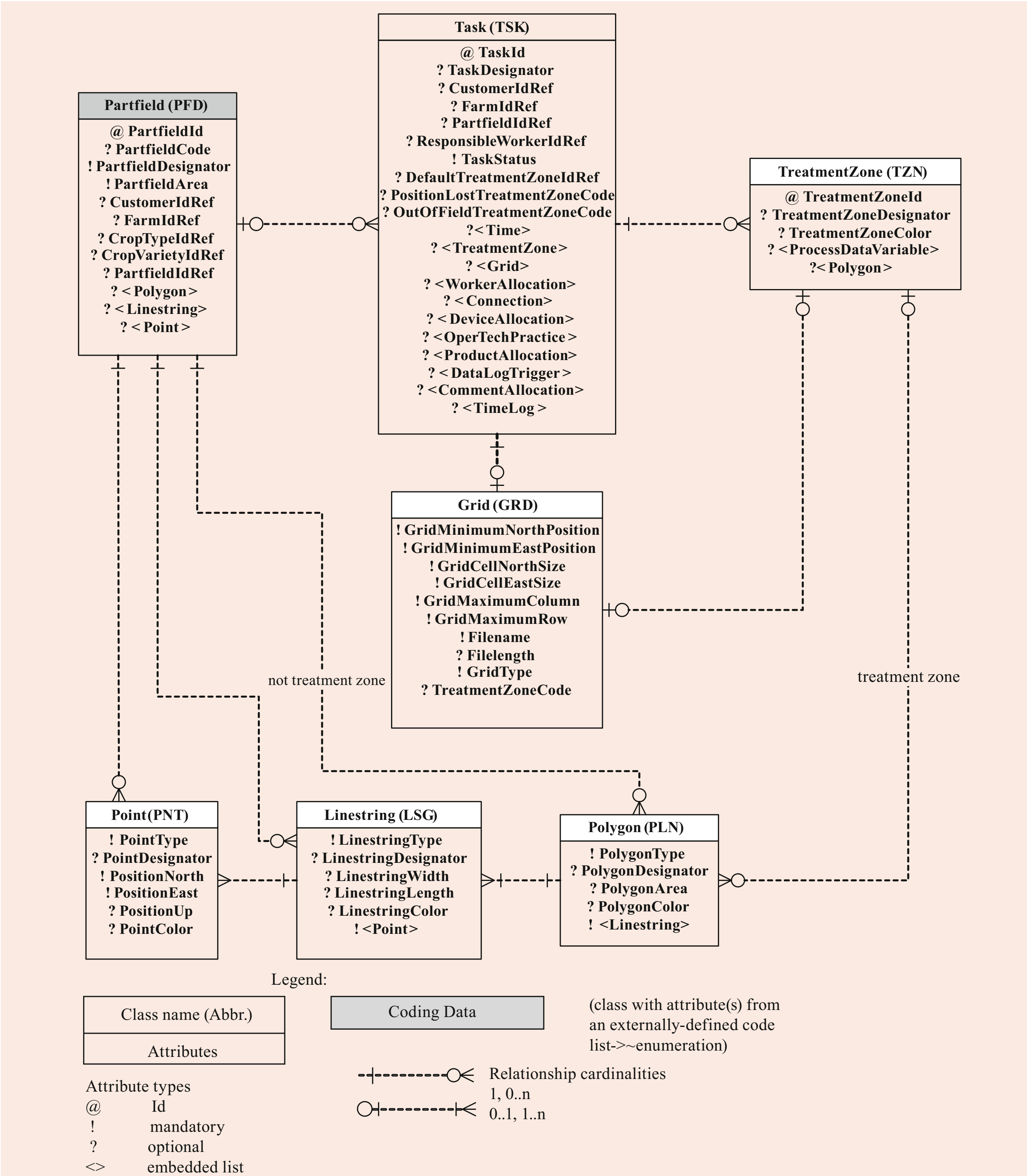


Fig. 24.16 Extract of ISOBUS XML elements relationship diagram showing spatial data components. (After ISO 11783-10, annex C)

Spatial Data Infrastructure for Agriculture

Geodata are increasingly being made available in digital form over the Internet, not only to stationary desktop personal computers (PCs) but also to mobile devices. With the construction of *spatial data infrastructures* (SDI), which are based on international standards and allow interoperable use of geoinformation at every place and time, work processes in agriculture are also changing. Since *Nash* et al. [38] modeled and developed various SDI-based scenarios for precision farming, only a few implementation trials have been done.

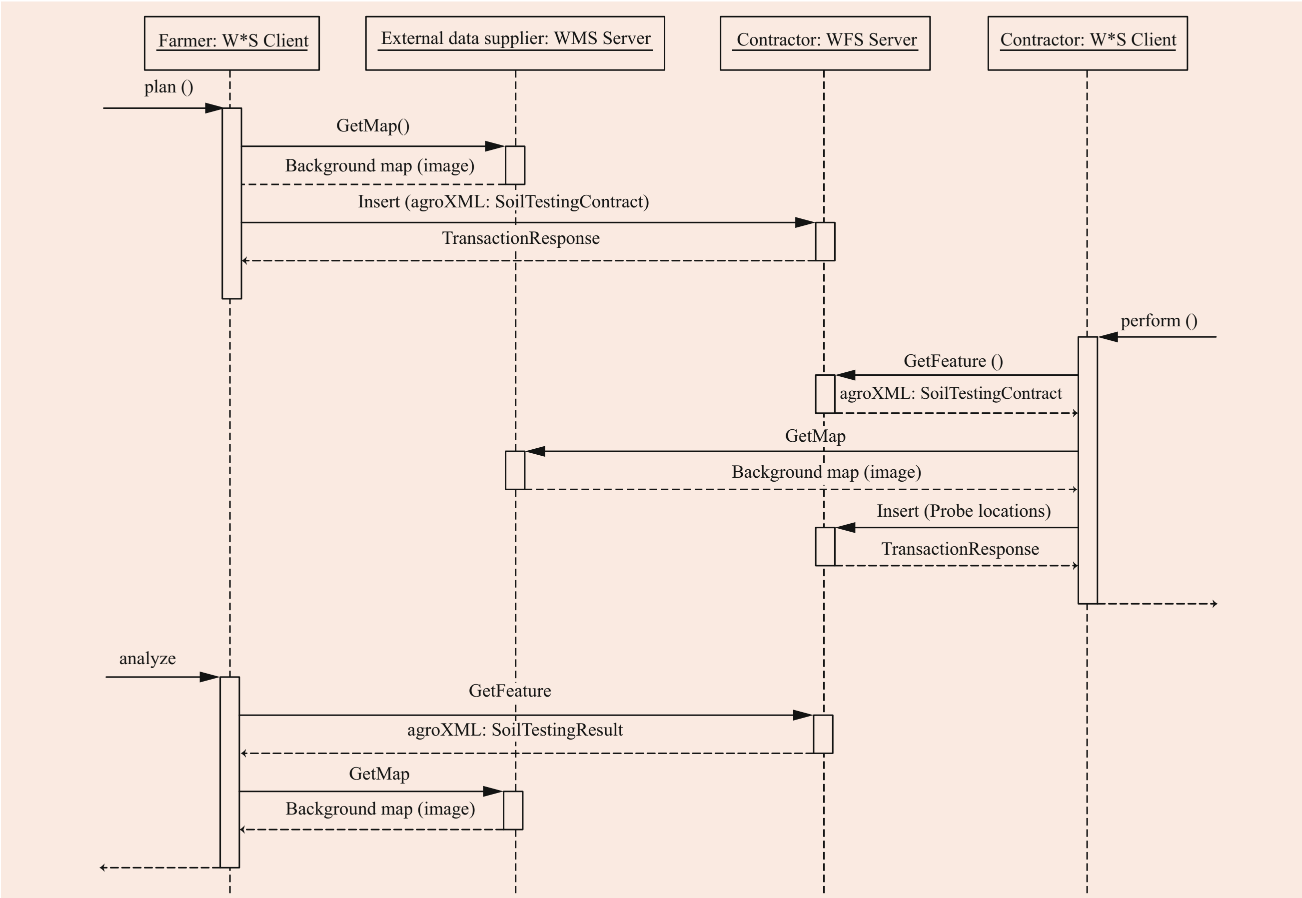


Fig. 24.17 Sequence diagram for soil testing with data transfers implemented using OGC interfaces and agriculture-specific data formats

It has to be stated that there is a lack of adoption of SDI-technologies in agriculture. Even a major use case like the spatial application for subsidies in the EU has not been implemented using a standardized SDI. The software clients offered from most authorities and FMIS vendors are not, or only partly, SDI-ready but offer web services in a proprietary way. The most common usage of SDI in agriculture is the usage of satellite or parcel images as background maps via WMS. These use cases are easy to implement and are not as complex as those proposed by *Nash* et al. [38]. However, there is an apparent need for farmers to benefit from the advantages of an SDI in agriculture. Beside the technical issues there seem to be an organizational barrier among the included parties.

On the technical side, it can be stated that noncomplex scenarios are straightforward to implement with OGC web services. The underlying information of an SDI is already available in most organizations and has to be prepared for usage within an SDI. Once captured and made available via a web-service interface, the data may be used in many different processes, particularly when these are implemented in the context of web services [7].

Fig. 24.17 shows a sequence diagram for soil sampling and testing, showing how the required information flows may be implemented using a modern spatial data infrastructure. Farmers periodically perform soil testing on their fields or employ a contractor to perform the testing, on the basis of which site-specific plans for tillage, sowing, or fertilization will be made. In order to define the probe locations, existing geodata such as soil maps or geological maps from government geological surveys and topographic and cadastral maps from governmental cadastral and survey agencies as external data providers are used. The samples taken are analyzed by a laboratory in an agricultural research and testing agency. The results of the analysis are communicated to the farmer, or the contractor, who then produces, e.g., a soil nutrient map.

One example presented by *Nash* et al. [5] is the calculation of the required nitrogen fertilization using Open Geospatial Consortium (OGC) Web Processing Service (WPS) interfaces and an opaque service chain. The total required nitrogen content may be estimated in a simple form by subtracting the available soil mineral nitrogen content from the amount of nitrogen contained in the previously harvested crop. The yield data may in future be collected by a harvest

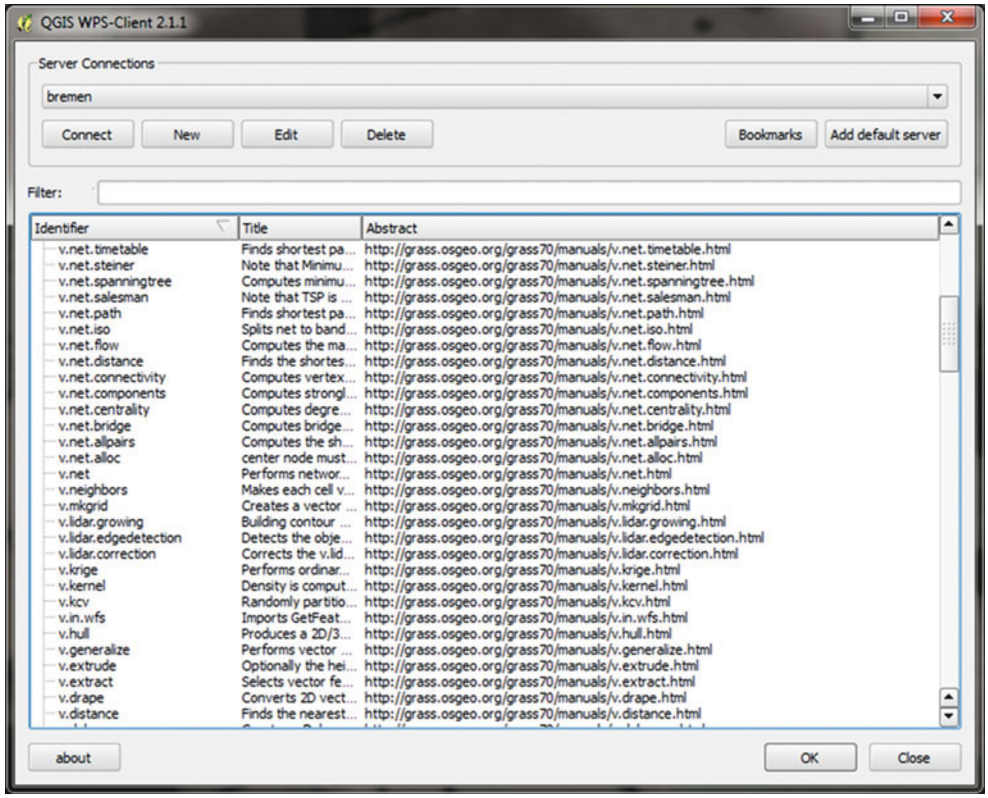


Fig. 24.18 QGIS-WPS-Client with a selection list of WPS processes

contractor and uploaded to an agricultural data warehouse, from which the data may be retrieved via a specialized OGC Web Feature Service (WFS) interface as part of an agricultural process data service [2].

Fig. 24.18 shows the provision of existing GIS processing algorithms (QGIS) remotely via WPS in WPS-aware client software.

Starting from noncomplex scenarios there are many use cases that are very helpful for farmers, service contractors, and authorities when they are delivered within an SDI. Fig. 24.19 shows how the required data processing may be implemented. In this scenario, the farmer does not have to store or manage any spatial data or perform any local processing locally. Given that problems related to data handling are frequently reported as being a reason for low uptake of precision farming [39, 40], the introduction of such a service-oriented architecture may be one means to increase acceptance of information-driven agriculture whilst also allowing access to powerful processing algorithms via mobile devices (e.g., from a smartphone in the field for on-the-spot decision-making) and to new scientific results, as improved models may be incorporated by simply swapping alternative services into the background chain whilst the interface remains unchanged.

One notable requirement for SDIs for agriculture is that current agricultural data formats are not based on Geographic Markup Language (GML), and so OGC services must be modified to manage these formats. In this way, a specialized SDI for agriculture may be developed.

Summary

The preceding paragraphs describe a modern high-technology agricultural system that uses various information sources in order to reach better decisions and, therefore, produce better products. Naturally, there are many practical problems and challenges in implementing such a system, which should be mentioned.

On the technical side, extremely complex systems are produced, which due to a lack of standardization, a large quantity of data, and multiple data exchange requirements, do not always function in union. Metadata are missing, and data security and archiving are as yet mainly unexplored themes. The costs for programs, contractors, and services are still high, and networking with other on-farm software does not always work.

For implementation in crop production, rules and algorithms for decision support and farm-specific functions are often missing. The time required for an individual analysis is immense, and only a small minority of farms have the required IT and GIS knowhow. Functionality and complexity compete with simple usage, meaning that there is a large requirement for training.

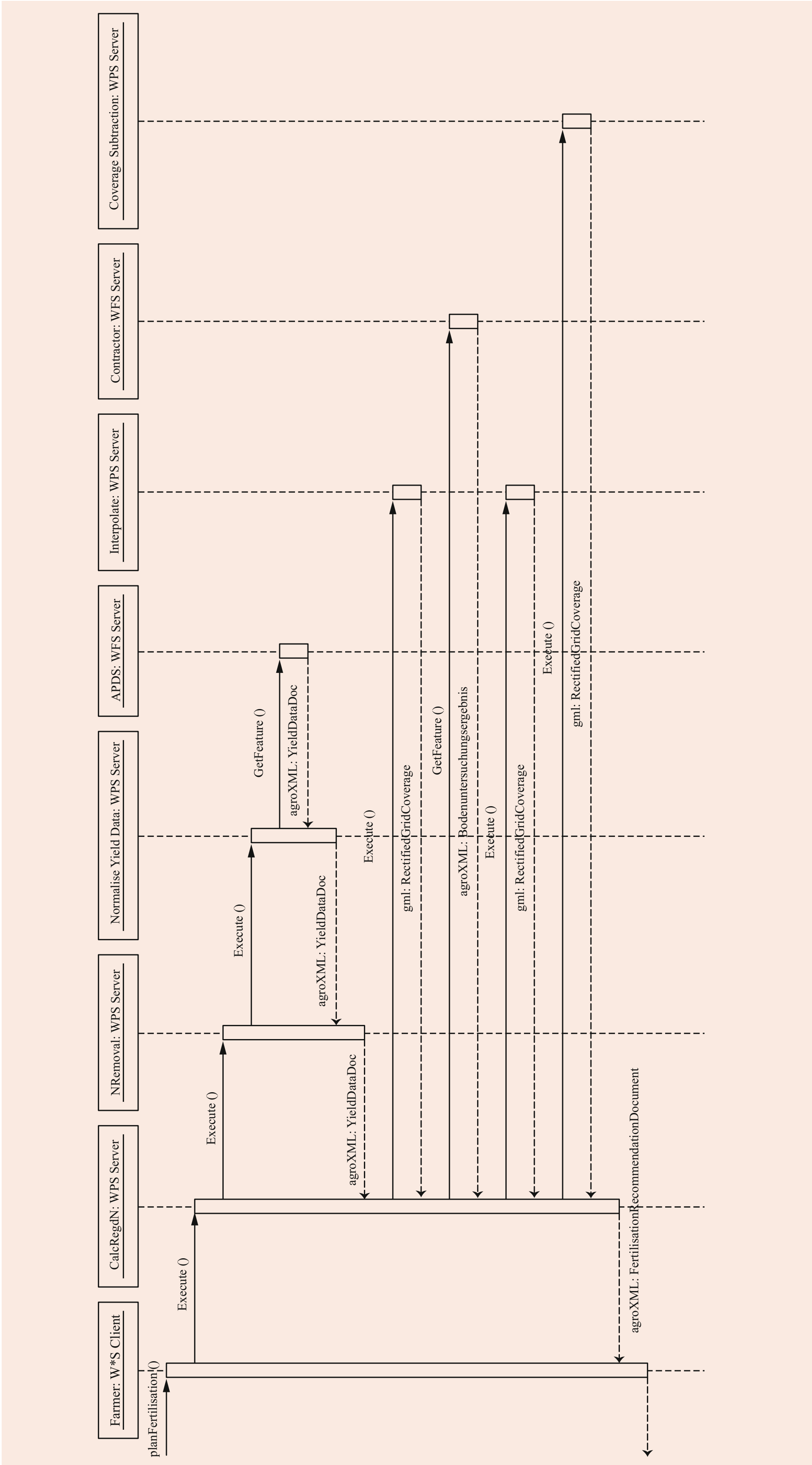


Fig. 24.19 Sequence diagram for N fertilization implemented with OGC services

24.4 GIS on the Farm of Tomorrow

24.4.1 Smart Farming

In recent years, the term "smart farming" or "smart agriculture" seems to have become increasingly more prominent. From the perspective of farmers, smart farming should provide added value in the form of decision support or the optimization of processes whenever and wherever needed. Some sources call this a next revolution in agriculture, after plant breeding and genetic engineering, influencing the agricultural world through the combination of ICT solutions, the Internet of Things (IoT), sensors and actuators, geopositioning systems, big data, unmanned aerial vehicles (UAVs), robotics, etc. Smart farming has the potential to support more productive and sustainable agriculture through a more precise and resource-efficient approach [41]. Accordingly, smart farming is closely linked to those technology areas that we described in this article:

- Management information system. Systems for collecting, processing, analyzing, storing, and communicating data in a form necessary for the execution of processes and functions in agriculture.
- Precision farming. Managing spatial and temporal variability to increase cost-effectiveness and reduce negative environmental impacts through optimized input. This includes decision support systems (DSS) for overall operations management to optimize revenue while conserving resources. The use of GNSS, drone aerial photography, and the latest generation of Sentinel satellite time-series images ensures the creation of high-resolution maps using a variety of factors (e.g., yield, terrain characteristics, topography, humus content, soil moisture, N status).
- Agricultural automation and robotics. The process of applying automation, robotics, and artificial intelligence to all levels of agricultural production, taking into account farmbots and farmdrones.

Smart farming applications not only target conventionally large farms but also have the potential to support family farms (small scale, specialized crops, rare species conservation) and organic farming. Furthermore, it allows for an accepted and transparent production in the sense of the European consumer. Smart farming also contributes to environmentally sound production, e.g., through efficient water use or optimized inventory management.

24.4.2 Trends in Farm Management

Farm management is currently undergoing many changes, which are driven by many different causes [42]. One common theme is the requirement for farmers to manage and exchange increasing volumes of information, much of which is spatially referenced. This, in turn, is driving various standardization initiatives in which spatial information and standards from the GI domain play a role. In this section, some current research trends relevant to GIScience will be presented briefly.

In many regions, changing economic conditions for farmers are leading to a decrease in small farms as smallholders or tenant farmers transfer their holdings to larger commercial enterprises, leading to an increase in farm sizes and increasing the distance between farm managers and the conditions in the fields. Alternatively, small-scale or hobby farmers in a region may cooperate by combining neighboring fields and/or sharing farm machinery in order to reduce the amount of low-yield field border regions and produce economies of scale, perhaps also through the use of larger machinery which could not be used in the individual land parcels. A further trend is the use of contractors to perform farm operations such as harvesting, reducing the need for the farmer to invest in specialized machinery.

In all of these cases, spatial information plays a role in the management; in larger farms, the lack of detailed local knowledge by the farm manager may mean that interpretation of data such as soil and yield mapping will play a more important role in the decision-making process. Where cooperative management is used, the amount of input and the yield from each of the contributing land parcels and, thus, the profit/loss for each farmer may be calculated based on spatially referenced data collected during field operations. Similarly, the exact region of operation is an important part of the contract information for a contractor, and the data collected during operation may be used to calculate the fee charged.

24.4.3 Standardization Activities in the Agricultural Information Domain

Until now, standardization in agricultural data transfer has concentrated on communication between field devices, e.g., the *ISOBUS* standards family. Transfer of data between software systems and between organizations has, with the exception of some limited proprietary and/or national-based standards (e.g., DAPLOS, EDI (electronic data exchange)), remained unstandardized and relied on bilateral agreements. Currently, multiple initiatives are attempting to produce XML-based transfer formats for agricultural information. The spatial properties of agricultural data are covered in these proto-standards to varying degrees but do not play a leading role. The use of GML, which would offer many advantages [43] including strongly object-oriented modeling, has become more popular through implementations of the European INSPIRE regulations.

During the course of the European project *GeoWebAgri*, the usage of GML and geo web services in the context of precision farming field operations is shown. However, the usage of these technologies in agriculture is not yet widespread, and the supporting community is still small.

The preceding activities may fit well for the standardization of specialized tasks within the agricultural domain. For the case of information exchange and integration with other domains (e.g., within the food chain and beyond), the usage of RDF (Resource Description Framework) can be a suitable supplement technology. Particularly for the GI-Sciences, the GeoSPARQL [44] language and vocabulary offers the representation and querying of geospatial information in RDF. GeoSPARQL extends the generic RDF query language SPARQL [45] with support for querying geospatial information. Thus, there is an opportunity to seamlessly integrate geospatial information with RDF-based information from other domains, including the exchange and processing of spatial and nonspatial rules in a standardized manner (OWL 2 [46], SWRL [47], RIF [48]). The introduction of these technologies may lead to a more efficient information exchange along the processing chain and implementations of crop-production standards with less burden for all partners (Sect. 24.4.4). In order to take advantage of RDF-based information the underlying GIS needs to support connections to RDF-databases (aka triple stores) such as Parliament [49] or Stardog [50]. Fig. 24.20 shows the usage of GeoSPARQL to check the presence of a geometry for an agricultural field.

24.4.4 Crop-Production Standards and Traceability

Farm management and crop-production standards are playing an increasingly important role in farm activities. Legal regulations control which fertilizers may be used and how and when they may be applied. Crop-production standards and associated product labels may be used to enforce good agricultural practice or conformance to a particular production system (e.g., organic farming). Finally, subsidy payments to farmers in the European Union are directly related to compliance with environmental measures through the cross-compliance scheme (Sect. 24.2). Each of these laws, regulations, and standards can be considered to define a set of individual rules that farmers must respect when planning and performing field operations. Currently, each farmer is likely to have to manually drawn up a personalized checklist against which operations are evaluated. This is complicated due to regional and local variations in rules, e.g., within nature and water protection areas, which may cover only part of a farm or part of a field; additional rules may be enforced.

Ways in which the process of compliance checking may be automated have already been researched [51]. Using a combination of machine-readable encoding of the actual rules together with metadata describing the regions and farmers to which they apply, the creation of a service-oriented architecture is proposed as a means to allow farm software to adapt dynamically to the local situation. However, the

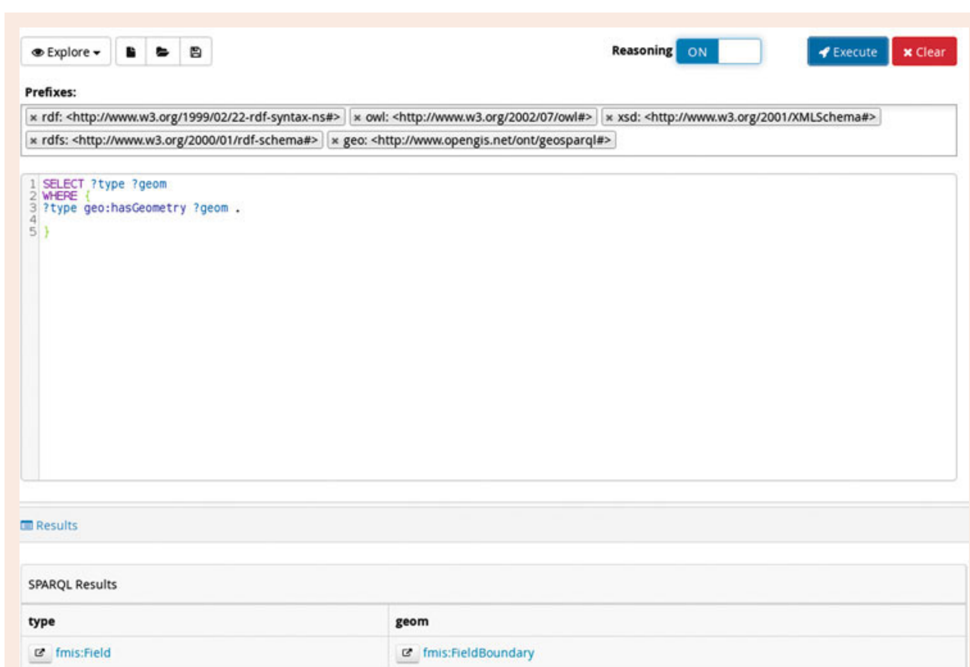


Fig. 24.20 Using GeoSPARQL to obtain a field boundary with Stardog

evaluation of rules requires large quantities of data, and nontrivial data processing, e.g., evaluating compliance to exclusion zones around water bodies requires the boundaries of these, potentially together with digital terrain models in order to calculate slopes. Where the broader effects of operations must also be considered, e.g., in regulating agricultural run-off, complex models and many geographic datasets may be demanded.

Food safety concerns, as well as consumer demand for regional produce and fair trade, are leading to increasing requirements for farmers to record all agronomic activities in detail, and for this information to accompany the actual produce through the processing chain so that in the case of contaminated food, it may be possible to swiftly trace the exact field of origin of the produce and, thereby, search this region for potential environmental sources of the contamination. In order to implement such a system efficiently, it is necessary to have standardized data transfer formats and procedures and a mechanism to link the information to the physical product, e.g., through barcodes or radiofrequency identification (RFID) tags that will be propagated along the chain. However, farmers are also concerned about the possibility of *transparent farming* and loss of a private sphere for both themselves and their business. Data protection and security mechanisms are, therefore, also an important component: all actors must only be able to see the relevant data, and ideally any access will require permission from the data owner.

24.4.5 Robotics

Two trends in the automation of extensive farm operations with robots may be observed:

1. Automation of existing large-scale machinery (of which autosteering may be considered a part), such that existing operational techniques may in the future be performed by large, automated vehicles with the human operator increasingly becoming an observer.
2. Use of fleets of small robots that allow use of new and novel techniques in crop management [52–55].

In both of these cases, spatial information plays an important part both in planning and in documenting operations; either an exact, spatially referenced plan must be prepared in advance and used to program the robotic operation, or a more general region of operation must be defined within which the robot may operate independently using sensor inputs to control the operation. Particularly in the latter case, it is necessary to document also exactly which operations the robot has performed in the field, including the location of each individual step. In cases where small, energy-limited robots are to be used, resource-aware positioning techniques developed for wireless geosensor networks may be necessary in place of GNSS.

24.5 Outlook

Due to legal regulations (IACS, cross compliance, traceability, quality management, etc.), GIS (and geo web services) and information-driven crop production are becoming normal tools in agriculture, which must be integrated into usual farm practices. It is not realistic to expect farmers to maintain multiple separate information systems, and so the farm GIS must be fully integrated into the typical record-keeping software in use on farms. Data and services provided via the Internet will to some extent reduce the role of the farm GIS in the future. Regional service providers will gradually have an even more important role. Until then, there is much research and development necessary (standards, interoperability, metadata, workflow optimization, etc.), which will maximize automation in the management and processing of agricultural geodata.

References

1. Pierce, F.J., Clay, D. (eds.): GIS Applications in Agriculture. CRC, Boca Raton, p. 224 (2007)
2. Steinberger, G., Rothmund, M., Auernhammer, H.: Mobile farm equipment as a data source in an agricultural service architecture. Comput. Electron. Agric. **65**(2), 238–246 (2009)
3. van Diepen, F., Charvat, K., Dittmann, C., Jezek, J., Martini, D., Wagner, R.: AgriXchange – European data exchange in agriculture. In: Proc. Workshop LPIS Appl. Qual., Sofia. EC Joint Research Centre, Ispra (2008)
4. Fountas, S., Wulfsohn, D., Blackmore, B.S., Jacobsen, H.L., Pedersen, S.M.: A model of decision-making and information flows for information-intensive agriculture. Agric. Syst. **87**(2), 192–210 (2006)
5. Nash, E., Korduan, P., Bill, R.: Applications of open geospatial web services in precision agriculture: A review. Precis. Agric. **10**(6), 546–560 (2009)
6. Bill, R.: Interoperable GIS-Infrastruktur im landwirtschaftlichen Betrieb. In: Dachverband wissenschaftlicher Gesellschaften der Agrar-, Forst-, Ernährungs-, Veterinär- und Umweltforschung (ed.) Geographische Informationssysteme in der Landwirtschaft und im ländlichen Raum. DAF, KTBL, Darmstadt (2004)
7. Hufnagel, J., Herbst, R., Jarfe, A., Werner, A.: Precision Farming – Analyse, Planung, Umsetzung in der Praxis. KTBL Schrift, vol. 419. Landwirtschaftsverlag, Müster (2004)
8. Mulla, D.J.: Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosyst. Eng. **114**, 358–371 (2013)
9. Henault, F., Brouant, B.: Accuracy of 10 guidance systems for farming under the same conditions at the same time. In: Proc. 2005 EFITA/WCCA Jt. Congr. IT Agric. pp. 687–694. EFITA, Vila Real (2005)
10. Grenzdörffer, G., Donath, C.: Generation and analysis of digital terrain models with parallel guidance systems for precision agri-

culture. In: Proc. 1st Int. Conf. Mach. Control Guid. pp. 141–150. ETH, Zürich (2008)
11. Peets, S., Mouazen, A.M., Blackburn, K., Kuang, B., Wiebensohn, J.: Methods and procedures for automatic collection and management of data acquired from on-the-go sensors with application to on-the-go soil sensors. Comput. Electron. Agric. **81**(0), 104–112 (2012). https://doi.org/10.1016/j.compag.2011.11.011
12. Kay, S., Milenov, P.: Status of the Implementation of LPIS in the EU Member States. EC Joint Research Centre, Ispra (2008)
13. Léo, O., Lemoine, G.: Land Parcel Identification Systems in the Frame of Regulation (EC) 1593/2000, Version 1.4. EC Joint Research Centre, Ispra (2001)
14. Werner, A., Jarfe, A.: Precision Agriculture. Herausforderung an Integrative Forschung, Entwicklung und Anwendung in der Praxis. KTBL Sonderveröff., vol. 038. Landwirtschaftsverlag, Münster (2002)
15. Gebbers, R., Adamchuk, V.I.: Precision agriculture and food security. Science **327**(5967), 828–831 (2010). https://doi.org/10.1126/science.1183899
16. Huhtala, A., Suhonen, K., Mäkelä, P., Hakojärvi, M., Ahokas, J.: Evaluation of instrumentation for cow positioning and tracking indoors. Biosyst. Eng. **96**(3), 399–405 (2007)
17. Ipema, A.H., Bleumer, E.J.B., Hogewerf, P.H., Lokhurst, C., de Mol, R.M., Janssen, H., van der Wal, T.: Recording tracking behavior of dairy cows with wireless technologies. In: Precision Livestock Farming '09 – Proc. 4th Eur. Conf. Precis. Livest. Farming, Wageningen, pp. 135–142. Wageningen Academic, Wageningen (2009)
18. Bishop-Hurley, G.J., Swain, D.L., Anderson, D.M., Sikka, P., Crossman, C., Corke, P.: Virtual fencing applications: Implementing and testing an automated cattle control system. Comput. Electron. Agric. **56**(1), 14–22 (2007)
19. Resnik, B., Bill, R.: Vermessungskunde für den Planungs-, Bau- und Umweltbereich. Wichmann, Heidelberg, p. 380 (2018)
20. Bill, R.: Grundlagen der Geo-Informationssysteme. Wichmann, Heidelberg (2016)
21. Fountas, M., Carli, G., Sorensen, C.G., Tsiropoulos, Z., Cavalaris, C., Vatsanidou, A., Liakos, B., Canavari, M., Wiebensohn, J., Tisserye, B.: Farm management information systems: Current situation and future perspectives. Comput. Electron. Agric. **115**, 40–50 (2015). https://doi.org/10.1016/j.compag.2015.05.011
22. Korduan, P., Grenzdörffer, G., Bill, R.: Informationsmanagement und Informationsbeschaffung in der modernen Landwirtschaft. In: Conf. Dig. 2nd Wismarer Wirtschaftsinformatiktage, pp. 282–291. Fachhochschule Wismar, Wismar (2000)
23. Corwin, D.L., Lesch, S.M.: Apparent soil electrical conductivity measurements in agriculture. Comput. Electron. Agric. **46**(1–3), 11–43 (2005). https://doi.org/10.1016/j.compag.2004.10.005
24. Herbst, R.: Bodenschätzung, geoelektrische Sondierung und pedostatistische Modellierung als Basis von digitalen Hof-Bodenkarten im Präzisen Landbau. Schriftenreihe des Instituts für Pflanzenernährung und Bodenkunde, vol. 60. Univ. Kiel, Kiel (2002)
25. Morellos, A., Pantazia, X.-E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebensohn, J., Bill, R., Mouazen, A.M.: Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. Biosyst. Eng. **152**, 104–116 (2016). https://doi.org/10.1016/j.biosystemseng.2016.04.018
26. Reichelt, M., Jürgens, C.: Adoption and future perspective of precision farming in Germany: Results of several survey among different agricultural target groups. Precis. Agric. **10**(1), 73–94 (2009)
27. Lee, D.H., Sudduth, K.A., Drummond, S.T., Chung, S.O., Myers, D.B.: Automated yield map delay identification using phase correlation methodology. Trans. ASABE **55**(3), 743–752 (2012)
28. Blackmore, S., Marshall, C.: Yield mapping: Errors and algorithms. In: Robert, P.C., Rust, R.M., Larsen, W.E. (eds.) Proc. 3rd Int. Conf. Precis. Agric. p. 403. ASA CSSA SSSA, Madison (1996)
29. Blackmore, S., Godwin, R.J., Fountas, S.: The analysis of spatial and temporal trends in yield map data over six years. Biosyst. Eng. **84**(4), 455–466 (2003)
30. Fountas, S., Ess, D.R., Sørensen, C.G., Hawkins, S.E., Pedersen, H.H., Blackmore, B.S., Lowenberg-Deboer, J.: Information sources in precision agriculture in Denmark and the USA. In: Stafford, J., Werner, A. (eds.) Proc. 4th Eur. Conf. Precis. Agric., Wageningen. Wageningen Academic, Wageningen (2003)
31. Schmidt, F.: Hochgenaue Digitale Geländemodelle – Untersuchungen zur Erstellung, Analyse und Anwendung in der Landwirtschaft (2003). http://www.geoinformatik.uni-rostock.de/schriften/schriften_einzel.asp?DiplID=131
32. Nugteren, A., Robert, P.: Usefulness and feasibility of high accuracy digital elevation models for precision management. In: Stafford, J. (ed.) Precision Agriculture '99 – Proc. 2nd Eur. Conf. Precis. Agric., Odense, pp. 561–569. Sheffield Academic, Sheffield (1999)
33. Constantin, D., Rehak, M., Akhtman, Y., Liebisch, F.: Detection of crop properties by means of hyperspectral remote sensing from a micro UAV. Bornimer Agrartech. Ber. **88**, 129–137 (2015)
34. Grenzdörffer, G.: Crop height determination with UAS point clouds. Int. Arch. Photogramm. Remote. Sens. Spatial Inf. Sci. **XL-1**, 135–140 (2014). https://doi.org/10.5194/isprsarchives-XL-1-135-2014
35. Whelan, B., McBratney, A.: Definition and interpretation of potential management zones in Australia. In: Solutions for a Better Environment – Proc. 11th Aust. Agron. Conf. Australian Society of Agronomy, Geelong (2003)
36. Watson, M., Lowenberg-DeBoer, J.: Who will benefit from GPS auto guidance in the corn belt? (2002). http://www.agriculture.purdue.edu/ssmc/Frames/WhoGPSAutoGuidanceCornBelt.htm
37. Niemann, H., Schwaiberger, R., Fröba, N.: Parallelfahrsysteme 67. KTBL, Darmstadt (2007). http://www.ktbl.de/
38. Nash, E., Bill, R., Bobert, J.: Anwendungsfallanalyse für den Einsatz von GDI-Technologien in Precision Farming. GIS Z. Geoinformat. **11**, 12–19 (2007)
39. Fountas, S., Blackmore, S., Ess, D., Hawkins, S., Blumhoff, G., Lowenberg-Deboer, J., Sørensen, C.G.: Farmer experience with precision agriculture in Denmark and the US eastern corn belt. Precis. Agric. **6**(2), 121–141 (2005)
40. McBratney, A., Whelan, B.: Future directions of precision agriculture. Precis. Agric. **6**(1), 7–23 (2005)
41. Walter, A., Finger, R., Huber, R., Buchmann, N.: Smart farming is key to developing sustainable agriculture. Proc. Natl. Acad. Sci. U. S. A. **114**(24), 6148–6150 (2017). https://doi.org/10.1073/pnas.1707462114
42. Charvat, K., Gnip, P.: Analysis of external drivers for farm management and their influences on farm management information systems. In: van Henten, E.J., Goense, D., Lokhorst, C. (eds.) Precision Agriculture '09 – Proc. 7th Eur. Conf. Precis. Agric., Wageningen. Wageningen Academic, Wageningen (2009)
43. Korduan, P., Nash, E.: Integration von ISO- und agroXML in GML, INFORMATIK 2005 Informatik LIVE! In: Cremers, A.B., Manthey, R., Martini, P., Steinhage, V. (eds.) Beiträge der 35. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 2005, Bonn, pp. 375–379. (2005)
44. OGC: GeoSPARQL – A geographic query language for RDF data (2011). http://www.opengeospatial.org/standards/geosparql
45. W3C: SPARQL 1.1 overview (2013). https://www.w3.org/TR/sparql11-overview/
46. W3C: OWL 2 Web ontology language document overview (second edition) (2012). https://www.w3.org/TR/owl2-overview/

24

47. W3C: SWRL: A semantic Web rule language combining OWL and RuleML (2004). https://www.w3.org/Submission/SWRL/
48. W3C: RIF overview (second edition) (2013). https://www.w3.org/TR/rif-overview/
49. . http://parliament.semwebcentral.org
50. Stardog: Website (2021). https://www.stardog.com
51. Nikkilä, R., Wiebensohn, J., Nash, E., Seilonen, I., Koskinen, K.: A service infrastructure for the representation, discovery, distribution and evaluation of agricultural production standards for automated compliance control. Comput. Electron. Agric. **80**, 80–88 (2012). https://doi.org/10.1016/j.compag.2011.10.011
52. Tillett, N.D., Hague, T., Marchant, J.A.: A robotic system for plant-scale husbandry. J. Agric. Eng. Res. **69**(2), 169–178 (1998)
53. Astrand, B., Baerveldt, A.J.: An agricultural mobile robot with vision-based perception for mechanical weed control. Auton. Robots **13**(1), 21–25 (2002)
54. van Henten, E.J., van Asselt, C.J., Blaauw, S.K., Govers, M.H., Hofstee, J.W., Jansen, R.M., Nieuwenhuizen, A.T., Speetjens, S.L., Stigter, J.D., van Straten, G., van Willigenburg, L.G.E.J.: WURking: A small sized autonomous robot for the farm of the future. In: van Henten, E.J., Goense, D., Lokhurst, C. (eds.) Precision Agriculture '09 – Proc. 7th Eur. Conf. Precis. Agric., Wageningen, pp. 833–840. Wageningen Academic, Wageningen (2009)
55. Emmi, L., Gonzalez-de-Soto, M., Pajares, G., Gonzalez-de-Santos, P.: New trends in robotics for agriculture: Integration and assessment of a real fleet of robots. Sci. World J. (2014). https://doi.org/10.1155/2014/404059

Further Reading

56. Noack, P.O., Muhr, T., Demmel, M., Stafford, J., Werner, A.: An algorithm for automatic detection and elimination of defective yield data. In: Proc. 4th Eur. Conf. Precis. Agric., vol. 1, pp. 445–450. (2003)
57. Zhang, C., Kovacs, J.M.: The application of small unmanned aerial systems for precision agriculture: A review. Precis. Agric. **13**, 693 (2012). https://doi.org/10.1007/s11119-012-9274-5
58. Sørensen, C.G., Rodias, E., Bochtis, D.: Auto-steering and controlled traffic farming – Route planning and economics. In: Pedersen, S., Lind, K. (eds.) Precision Agriculture: Technology and Economic Perspectives. Progress in Precision Agriculture, pp. 129–145. Springer, Cham (2017)
59. Nikkilä, R., Nash, E., Wiebensohn, J., Seilonen, I., Koskinen, K.: Spatial inference with an interchangeable rule format. Int. J. Geogr. Inf. Sci. **27**(6), 1210–1226 (2013)

Ralf Bill Professor Ralf Bill studied geodesy in Mainz, Berlin, and Karlsruhe and worked as a scientist at the universities of Karlsruhe and Stuttgart, as well as a development engineer in Switzerland. Since 1994, he has been Professor of Geodesy and Geoinformatics at the University of Rostock and since then has also been involved in research on precision farming.

Edward Nash Edward Nash studied Mapping Information Science and completed a PhD in Geomatics at the University of Newcastle-upon-Tyne. Subsequently he worked as a Research Associate at Rostock University participating in national and European projects on precision farming, focussing on applications of spatial data infrastructures and web services. He is currently a developer of geospatial technologies, mainly for state government use.

Görres Grenzdörffer Görres Grenzdörffer studied geography at Tübingen University. After his diploma in 1994, he became a lecturer at Rostock University. In 2001, he received a PhD in Geoinformatics. Currently, Dr Grenzdörffer is Senior Lecturer at the Chair for Geodesy and Geoinformatics at Rostock University. He leads the group for remote sensing and precision farming. His research focuses on the use of drones for ecological and agricultural purposes.

Jens Wiebensohn Jens Wiebensohn finished his studies of agricultural ecology at Rostock University with an MSc focusing on GIS technologies in modern agriculture. As a Research Associate at Rostock University he participated in several European projects related to precision agriculture. Jens is currently working as a farm manager on an arable farm in Northern Germany.

Geographic Information Systems in Defense

25

Gerhard Joos

Contents

Abstract

This chapter is written from the perspective of the defense geospatial policy of nations with a *Western* orientation. This implies (but is not restricted to) the North Atlantic Treaty Organization (NATO) and partner countries. Whereas general geospatial principles apply independently of any country's affiliation to any alliance, there are standards and conventions discussed in this chapter that may not apply to other countries. Concepts of the United Nations Department of Peacekeeping Operations (UN DPKO) are also considered.

Command and control (C2) systems support the decision-making processes of commanders at different levels in the chain of command. C2 systems rely heavily on accurate, current, and complete area-wide geospatial data coverage of the area of interest. After introducing some basic terminology for military operations, including typical scenarios illustrating where and how geospatial data are needed in the military context, this chapter describes the concepts of the common operational picture and the recognized environmental picture, which form the basis for core geospatial services and especially functional services, i. e., services that rely on information with a geospatial component.

G. Joos (✉)
Department of Geoinformatics, Munich University of Applied Sciences
Munich, Germany
e-mail: gerhard.joos@hm.edu

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_25

As geospatial information is key to all other services within a military command, geospatial services are considered to be core services and are consequently called *core geospatial services*. These comprise the services for the capture, integration, transformation, and dissemination of geospatial base data together with the provision of metadata, as well as a service to enable users to search for and retrieve the data and processing functionalities they need for their functional area system. In this chapter, the importance of geographic information systems (GIS) is demonstrated for a selection of functional systems such as C2, logistics, intelligence, surveillance and reconnaissance, air-traffic control, the airborne warning and control system (AWACS), confidence-building measures and disarmament, electronic warfare analysis systems, embedded systems, and training and simulation. Military standards supporting interoperability and a selection of geospatial data products that are typically applied for military applications complete this chapter on GIS in defense.

Keywords

GIS in defense · operational picture · command and control systems · core geospatial services · ISR – intelligence, surveillance & reconnaissance · geospatial military standards · military geodata

25.1 Background and History

Military operations and the need for technological superiority to protect against potential hostile attacks have always been a driving force in the development of technology and, in particular, cartography. For centuries, cartography has been crucial for commanders, as it can support the decision-making process and give answers regarding the spatial dimension to questions concerning the six Ws of battlefield management: who, what, why, how, where, and when. Thus, with the emergence of electronic weapon systems, the digitalization of warfare, and the need to accelerate decision support, geospatial information systems have become an indispensable part of military operations. This chapter will give an overview of application areas, data product specifications, and military standards for such systems in the defense field.

In this sense, command, control, communication, computers, intelligence, surveillance, and reconnaissance (C4ISR) is one of the most important fields of application for geospatial information. This chapter gives an insight into a few areas that demonstrate the importance of geospatial information for the safety of civilians in peacetime and for troops in missions.

To provide an understanding of military jargon, the two important terms *joint* and *combined* are introduced. The term *joint* implies a combination of the branches of military forces, i. e., the army, air force, and navy. Some countries do not have all three branches; in particular, countries without access to the sea usually do not have military vessels and so do not need a navy. Other countries, such as the USA, consider the coast guard to be an additional branch. *Combined* operations involve the cooperation of troops from different countries. Combined operations are therefore multinational operations.

Military operations are executed at different levels of command: strategic, operational, and tactical. The strategic level refers to long-term and long-range planning of operations of military forces at a rather political level, and therefore usually requires small-scale maps or low-resolution geospatial data with a global, regional, or national area of interest. The operational level refers to operations within an area of interest. More detailed geospatial information is required at the operational level. The tactical level involves executing the strategic and operational objectives. Combat troops can be seen as the smallest tactical units. They require the most detailed, accurate, and up-to-date geospatial information.

The strategic, operational, and tactical levels are also reflected in the levels of commands. Strategic planning is a political responsibility and is therefore conducted in parliaments, ministries, or the highest headquarters (HQ) of combined operations. Operational and tactical planning takes place in subordinate echelons of a hierarchical command structure. Headquarters at the operational level must act as brokers between strategists and executives. They can be structured in a hierarchy of several HQs and commands. Commands for the tactical level are deployed in the theater of operations.

The different tasks involved in a command are divided into directorates with specific responsibilities. Table 25.1 shows the clustering and identification of a J-code organization by numbering from 1 to 9, as used in many countries. The responsibilities may vary from country to country.

Table 25.1 Directorates in a military organization and their responsibilities [1]

Directorate	Responsibilities of the joint staff
J-1	Manpower and personnel
J-2	Intelligence
J-3	Operations
J-4	Logistics
J-5	Strategy, plans, and policy
J-6	Command, control, communications, and cyber defense
J-7	Training and education
J-8	Resource management and financial support
J-9	Civil–military operations directorate/interagency operations

Most directorates require geospatial data as well as systems to analyze data and visualize the results of the analysis in reports and briefings to their commanders. Since they all have a need for base data, one directorate is responsible for providing the others with geospatial base data, such as satellite or airborne orthorectified and georeferenced imagery, scanned maps, topographic data, political boundaries, terrain data, weather data, lines of communication, possibly three-dimensional (3-D) models, and other relevant overlay data. To operate on the basis of the same topographic and geospatial intelligence information, all the directorates in a command and all the ordinate and subordinate commands need to be supplied with the same geospatial base data.

Geospatial information systems are used to support all phases of military operations, such as:

- Simulation
- Mission planning
- Mission rehearsal
- Command and control
- Impact analysis
- Post-mission analysis.

25.2 Scenarios

Operational scenarios are useful to demonstrate the usage and need for geospatial data and geospatial functionality in operational use cases. By using the language of the end user, they help the user to express their need and geospatial specialists to understand the technological requirements. The Defence Geospatial Information Working Group (DGIWG), a standardization body for GIS-related aspects in the military community, has compiled a number of operations that are published on their website [2].

25.2.1 Coalition War Fighting Operation

If a country carries out unacceptable aggressive activities against another country, the United Nations (UN) could issue resolutions [3] against the aggressor. If the country under attack is part of NATO, it may invoke Article 5 of the Treaty, prompting the establishment of a coalition force with options for military deployment. This collective self-defense force may include both armored and mechanized ground forces that are able to mount land-based, amphibious, and airborne assaults using both conventional troops and special force formations.

The coalition war fighting operation scenario covers a wide range of geospatial support activities at the strategic, operational, and tactical levels. Coalition members rehearse all kinds of scenarios to gain common experience in this type of operation. For this reason, there is a high level of operational preparedness, with many off-the-shelf and customized GIS products and geospatial services being readily applicable and interoperable.

25.2.2 Coalition Peacekeeping Operation

When there is fighting between warring factions, coalition forces are stationed to assist in the development of a stable government, in educating law enforcement personnel, and in rebuilding the local economy and local critical infrastructure. This coalition force has strong peace enforcement capabilities, including armored formations and airmobile troops. It also has significant civil affairs, engineering support, and logistic service capabilities at its disposal. The command of the force is usually managed by a lead nation, with coalition troops being integrated into one chain of command.

The operation covers a wide range of geospatial support activities, primarily at the operational level of planning. Geospatial support is provided to both military activities and civil engineering activities (CIMIC). Since multiple coalition members are involved, interoperability is a major requirement. It can be assumed that the members already have common experience and therefore there is a high level of operational preparedness supported by off-the-shelf products and services.

25.2.3 Counter-Insurgency Operation

In this scenario, a group of insurgents is based in a region external to any coalition member territory. Analysis from intelligence sources allows a counter-insurgency (COIN) operation to be mounted by one nation, but with cooperation with other states to ensure that all national and coalition COIN forces have the same information and support they need to mount successful measures against the insurgents.

Member nations who are not involved with undertaking the counter-insurgency operation may contribute geospatial support from their national sites rather than contributing in-theatre. This reach-back support predominantly takes the form of the exchange and integration of intelligence information. The challenge in this scenario is to achieve semantic interoperability by bringing together all available information in a common data model.

25.2.4 Noncombatant Evacuation Operation

An unstable political situation in a country with threats or emergency circumstances may cause a group of states to decide to evacuate nationals from the country of concern. This effort needs to be coordinated between the involved nations. One nation needs to take overall responsibility for coordinating the evacuation arrangements, with a variety of nations offering support. Before actively sending troops to conduct or supervise the evacuation, the people in the country need to be supported with current geospatial information, in particular regarding threat and no-go areas. Assembly points and evacuation corridors need to be identified and communicated. Military forces for this operation may include support helicopters, defensive troops, and either a naval/amphibious force or troops manning facilities in a neighboring stable country.

Geospatial support comes from nations who are members and nonmembers of a formal coalition. Procedures for geospatial support collaboration exist for members of coalitions but not for nonmembers. The interoperability of the geospatial products and services offered may be uncertain. Nevertheless, this information is very important for planning.

25.2.5 United Nations Humanitarian Aid Operation

Severe weather conditions or a natural catastrophe in a region lead to a state of disaster being declared by the local government, prompting foreign nations to offer assistance, e. g., within a UN humanitarian aid program. The military forces deployed may include support helicopters, medical units, engineering capabilities, and logistic support.

To identify the magnitude of need and ways to reach the affected areas in order to coordinate the relief operation, up-to-date geospatial data are required. Depending on the type of disaster, there will either be a significant amount of lead time in preparing for these operations, or, with natural catastrophes such as earthquakes, floods, and tsunamis, there is no preparation time at all. Main source of information are satellite imagery captured before and after the desaster. Change detection functionality is required.

25.2.6 Coalition Sanctions Enforcement Operation

As part of diplomatic efforts, UN sanctions may be applied against the corrupt government of a country. Coalition forces may deploy in order to enforce UN resolutions, as sanction violators may use sea or land routes to move contraband goods into and out of the country. The military forces deployed by the coalition will include strong intelligence-gathering capabilities supported by JISR to monitor forces in border areas. They will include forces that can deploy rapidly to remote areas where illegal activities are detected.

The lead nation will provide geospatial support to the central command. Given the nature of the military operation, nations are brought together from an existing coalition that has predetermined procedures for geospatial support. A moderate amount of time is available to prepare and plan for the operations. Given the long period (potentially one or more years) for which support must be provided, there is the opportunity to refine coalition procedures. The operational tempo for geospatial support is generally low, but there are sporadic periods of high activity when sanction-busting activities are detected. For a rapid response, in-theater capabilities are critical.

25.3 Situational Awareness

For effective decision-making, it is crucial that the commander in charge has a comprehensive overview of the situation. The common operational picture (COP) and the recognized environmental picture (REP) provide this overview information in a map representation at different levels of detail depending on the strategic, operational, or tactical command level.

25.3.1 Common Operational Picture

The common operational picture (COP) provides the integrated capability to receive, correlate, and display a common picture, including planning applications and theater-generated geospatial information overlays that may include the locations of friendly, hostile, and neutral units, assets, and reference points [4]. In principle, the COP consists of base maps generated from either imagery or rendered vector data together with different overlay information that give the user the most comprehensive picture of the current situation in the area of interest. The overlay data can be incidents, the distribution of friendly and hostile assets, civilian infrastructure, or any other georeferenced features.

For military assets, a language of tactical symbols has been developed. There are standards that define base tactical symbols and rules for how those base tactical symbols are composed and annotated. At a large-scale tactical level, single squads need to be represented, and at smaller scales the symbols need to be aggregated and replaced by symbols representing the next highest echelon in the command chain. From lowest level to a whole theatre this hierarchy of echelons applies: squad – section – platoon – company – battalion – regiment – brigade – division – corps – army – army group – theater or region (depending on the nation and the branch of service) [1].

The NATO standard for military map symbols is Allied Procedures Publication APP-6 *Military Symbols for Land Based Systems*. The NATO Standardization Agreement that describes APP-6 is STANAG 2019. The US derivative of APP-6 is known as MIL-STD2525 *Common Warfighting Symbology* [5]. These standards have evolved through several releases due to increasing requirements.

Table 25.2 Color coding of operational symbology [4]

Affiliation	Hand drawn	Computer generated
Friend, assumed friend	Blue	Cyan
Unknown, pending	Yellow	Yellow
Neutral	Green	Green
Enemy, suspect, joker, faker	Red	Red

The symbology standard is already in its fourth generation, as indicated by the letter D in APP-6D and MIL-STD2525D. New to this edition are symbols for cyber warfare, as more and more attacks against the infrastructure of a country are hostile acts conducted in cyberspace by governmental organizations. In addition to its virtual address, each component of a globally spreading computer network has a geospatial location, as each component manifests in hardware located at a well-defined location. Another new group of military symbols are concerned with assets orbiting the Earth or supporting space weapons.

The Department of Peacekeeping Operations of the United Nations also has a standard, which is similar to the NATO standard [6]. These standards are not all-inclusive, so other portrayal standards have to be applied, especially for base data, since military assets need to be based on background maps with a standardized symbology.

The APP-6 standard provides a common operational symbology along with details regarding their display and plotting to ensure the compatibility and—to the greatest extent possible—the interoperability of NATO land component command, control, communications, computer, and intelligence (C4I) systems for development, operations, and training. APP-6 addresses the efficient transmission of symbology information through the use of a standard methodology for symbol hierarchy, information taxonomy, and symbol identifiers.

The color coding used for the symbols is consistent across all standards. MIL-STD2525 distinguishes between colors hand-drawn on maps and the colors used for computer-generated symbols (Table 25.2).

How echelon indicators look and how they are attached to the top of the asset icon are illustrated in Fig. 25.1. The icon for the actual location of an asset is drawn with a solid line, and a dashed icon box is used for the projected location.

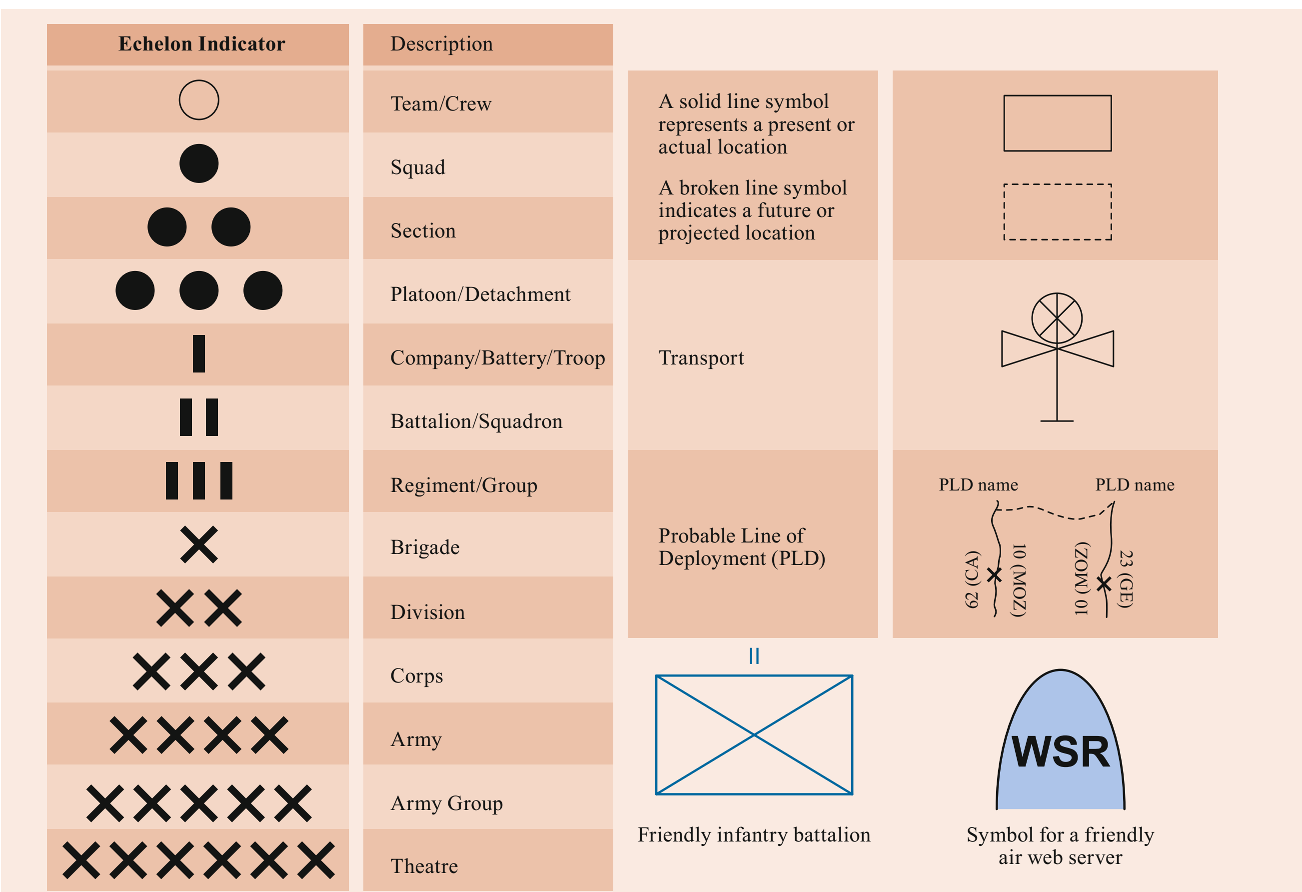


Fig. 25.1 Base symbols for the indicator of unit size (echelon indicator), the present and proposed locations, the symbol of a transportation aircraft, command and control operational symbols for lines of deployment, and a combined symbol used in NATO for a friendly infantry battalion and a friendly air web server (from MIL-STD 2525D [4])

The semantic of the asset is placed in the icon box. Annotations provide identifiers and additional description of the asset. Different line styles either indicate movement or represent line symbols such as boundaries.

There is also a requirement for geospatial portrayal under different environmental conditions, e. g., at nighttime, for small displays in handheld devices, or for onboard systems in military aircraft.

25.3.2 Recognized Environmental Picture

NATO defines the recognized environmental picture (REP) as the controlled information base for the combination of geospatial, meteorological, and oceanographic (GEOMETOC) data [7]. The REP is characterized by highly dynamic data, including actual, predicted, and past weather data.

A more holistic understanding of REP includes all weather-related derived geospatial information, including cross-country mobility based on soil and weather conditions or wind-dependent helicopter landing sites, or atmospheric-related line-of-sight analyses.

Weather prediction requires topographic and elevation data with a time series of densely distributed meteorological sensor data in combination with Earth observation satellite data and radio detection and ranging (radar) scatter images. Especially in mountainous areas, local weather trends require sophisticated algorithms and high-capacity computing power to calculate prediction models. Inaccuracies in any input data and in the prediction models lead to an unpredictable deviation of the forecast. For this reason, the REP represents a fusion service where the spatially portrayed outcome of weather prediction calculated on one or more supercomputers is integrated with geospatial base data.

Weather influences cross-country mobility and ground visibility and is therefore especially important for operation planning and simulation. Heavy winds influence whether it is possible to fly helicopters and manned or unmanned aircraft and navigate vessels. In this sense, the REP is an important component of command and control systems.

25.4 Network-Centric Warfare

Dissemination of real-time information and communication are crucial to the decision-making process. Therefore, all systems are connected via networks throughout the different command levels and directorates. The network usually complies with Internet technology standards, although it is physically separated from the web for security reasons. In this respect, national or multinational organizations have to provide all of the necessary services, ranging from domain name servers to the spatial data infrastructure for this network. The advantage of this is that there is no unauthorized access to the information on the net, so it is not possible to attack and manipulate or sabotage it unless someone is physically at a computer connected to the network.

Geospatial services need to provide the base data to all other services in a timely manner. Figure 25.2 shows a simplified view of the requirement for interoperability in a network-centric environment.

The common operational picture (COP) and the recognized environmental picture (REP) provide the situational context for military operations [8]. They are fed by geospatial base data that are either captured as vector data or as gridded coverages. Base data can be corrected and extended by geospatial data captured in the area of interest by a deployed GIS. Reconnaissance and surveillance data from different kinds of sensors are used for situational awareness and as source data for geospatial vector data. The georeference parameters and color models of raw imagery data usually do not fulfill the required data quality acceptance levels for absolute positional and thematic accuracy, so geometric and radiometric corrections are necessary. Meteorological information is required to feed the REP.

As the analysis of geospatial data is a requirement across the entire user community, it needs to be provided as a centralized functionality accessible to the user group that needs to know the analytical results. The analysis of geospatial data—also called geoprocessing—can lead to very sensitive information, as elevated interest in a certain area may indicate plans for upcoming military operations there. Knowledge of the areas in which a particular analysis is conducted can help to reveal the need for a potential action. Typical geoprocessing analyses for defense include those for cross-country mobility, potential helicopter landing sites (especially for medical or other evacuation operations), and visibility. The results of these analyses are influenced by the meteorological conditions.

All applications, or functional services as they are also called, require geospatial base data as either background imagery, a scanned or rendered map, or as features for further analysis. These applications may just require base data or a combination of base data with appropriate overlays. As weapon systems getting more and more equipped with sensors for geospatial information and communication capabilities their embedded computing power e.g. for navigation or targeting needs to be interoperable with C2-systems at tactical level. This enables weapon systems to become autonomous and to adjust for actual conditions. A selection of applications in the defense domain will be presented later in the chapter.

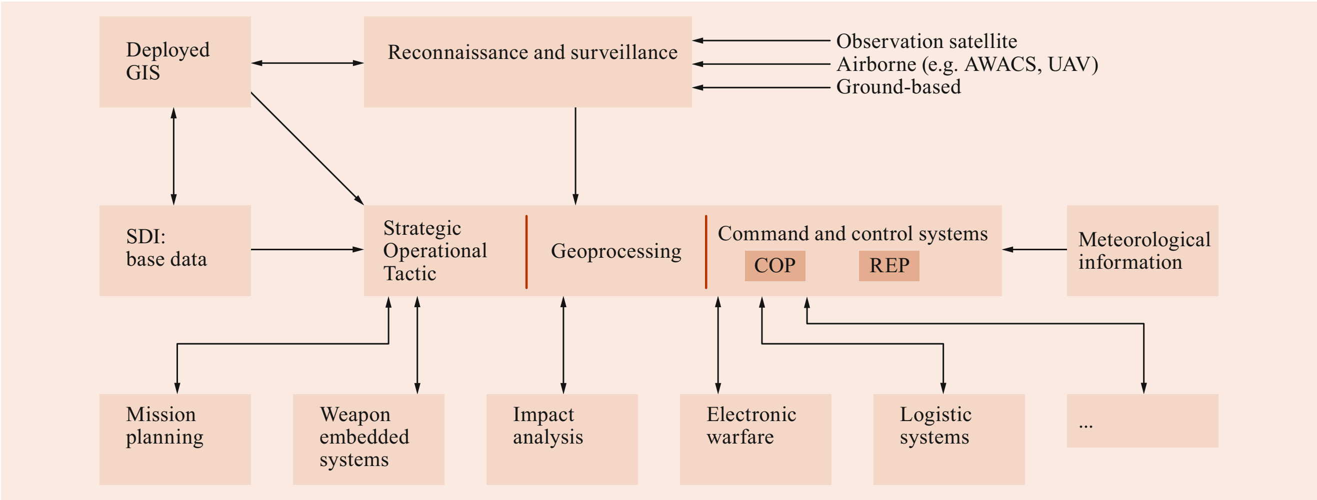


Fig. 25.2 Components and dependencies in a network-centric GIS architecture

25.5 Core Services

NATO categorizes services as being either *core* or *functional*. Core services are capabilities that are applicable to all operational users, such as document management, office automation, or geographic information systems. Functional services are only relevant and applicable for a subgroup of personnel in the command structure.

25.5.1 NATO Core GIS

Providing geospatial base data is a core service, since geospatial data is used in most functional services. The spatial data infrastructure (SDI) for military-relevant base data in the classified networks of NATO is called the NATO Core GIS. It consists of various components that interoperate via Open Geospatial Consortium (OGC) GeoWeb Service interfaces.

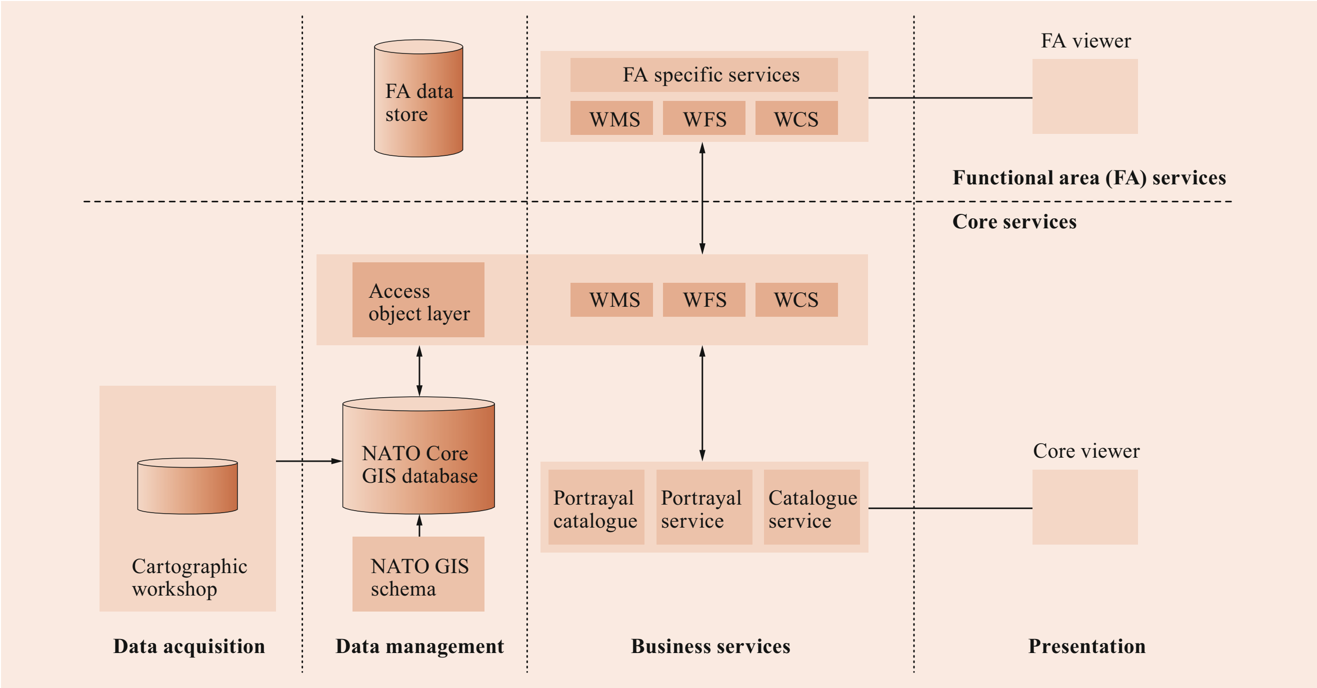


Fig. 25.3 NATO Core GIS architecture (simplified after [9])

The NATO Core GIS architecture is shown in Fig. 25.3. Designated base data are stored in the core GIS database. Functional services can discover available data by searching the catalog, and they can access the data via well-known OGC service interfaces. The symbols and symbology rules are maintained in a portrayal catalog. This includes symbols for base data and for the military-relevant tactical symbols. Rendering is done by a portrayal service. The core viewer can be used to visualize the geospatial data in the core GIS database and it provides some lightweight functionality such as distance or area measurements. More computationally intensive analyses are performed on the server and invoked via the Web Processing Service (WPS) interface. This avoids the need for the client to download huge amounts of geodata, which they may not be capable of. The portrayal may be visualized as a 2-D map or a 3-D perspective view. Functional services usually come with their own geospatial viewer embedded in the business logic of the functional application. Since the architecture is based on reusable components, the services can also utilize the core viewer.

The cartographic workshop, which is part of the NATO Core GIS architecture, is used to preprocess data and for small data production tasks. It consists of heavyweight GIS functionality to carry out quality assessments and to migrate data from any source (and in an arbitrary format) into the database. Every data unit must be associated with appropriate metadata that feed into the catalog service to enable users to find relevant datasets.

25.6 Functional Area Services

Decision-making, planning, simulation, and other applications for national defense require geospatial information. The following sections describe use cases for applications of geospatial data and geoanalytical tools in the field of defense. These applications show commonality in the handling of items that are spatially referenced to the Earth's surface. They therefore need to have some capability to visualize the locations of these items on a map or in a 3-D scene. Geospatial base data provide the reference to bring the items in prospect to the surrounding topographic features as we perceive our environment in the context of other features. For example to understand the threat originating from an improvised explosive device (IED) not only the location is required, but also the assets in the vicinity, e.g. hospitals, embassies, or other features of interest. The amount of integration of the functional geospatial items with the base data that is performed depends on the application. In most cases, the geospatial base data simply provide background information; they are not actually fused with the subject-specific data. In these cases, functional services typically access the NATO Core GIS via the Web Map Server (WMS) interface. If the base data are required for any kind of analysis, such as terrain analysis or routing through a node-edge network that is not part of the functional service, geospatial data need to be retrieved in an appropriate format that allows further analysis. In this case, the client application would need to retrieve geospatial data via either a Web Feature Service (WFS) or a Web Coverage Service (WCS), depending on the nature of the data. Alternatively, the analyses can be performed by the server and the results can be provided to the user via a Web Processing Service interface (WPS).

25.6.1 Command and Control

Whenever an operation includes several units that are spatially distributed, those units will need to be coordinated and directed. In the military context, there is a chain of command that reflects the hierarchy of units and commanders. The size of the area of operation usually increases up the hierarchy. In order to maintain an overview and full control, commanders need support from a system that provides them with situation awareness, allows analysis of the impacts of different scenarios to aid decision-making, and permits decisions to be transferred as orders down the chain of command, with further refinement performed at each echelon level. Systems that support this kind of decision-making are called command and control (C2) systems.

C2 has been defined by NATO as military function 01 [10]:

> The Organization, Process, Procedures and Systems necessary to allow timely political and military decision making and to enable military commanders to direct and control military forces.

C2 systems are further defined in NATO documents as including headquarters facilities, communications, information systems, and sensors and warning installations. Following this definition, most military information systems would fall into the C2 category. For clarity, this chapter treats other functional services in separate sections. In this respect, C2 systems find application in many areas of military activity, which may include search and rescue, combat operations, peacekeeping, or civil–military cooperation (CIMIC). Each area has different requirements for geospatial data, analysis functionality, documentation, or orders. For this reason, there are several C2 systems. It is the aim nowadays to combine highly specialized C2 systems into more integrated and joint systems.

The COP or REP is used to provide situational awareness. Cross-country mobility might be a requirement for some kinds of operations. Influential factors for cross-country mobility are the terrain and topographic information, and possibly soil data. Based on a combination of these, and taking current or forecasted weather data into account, a system can make a go/no-go decision as to whether different categories of vehicles should cross a certain area that is not accessible via the road network. Together with area-of-sight analysis and possible electromagnetic shadow areas, the system can support tactical operations. Different scenarios can be examined to weigh up options. The final decision is up to the commander, who then gives orders to the subordinate units.

In a network-centric environment, orders can also be generated and transmitted from the C2 system to clients. Orders also have a spatial dimension, e. g., the order may be to move from location A to location B. The client then needs to be able to read the order and combine it with the geospatial data stored on the client side. Since network connections in the field are usually very slow (if they exist at all), the amount of transmitted data needs to be minimized, which is why bulk geospatial data with less flux should reside on the client to free up bandwidth for the transmission of overlay data and orders.

There is a standardization committee that deals with the interoperability of C2 systems. It is called the Multilateral Interoperability Program (MIP) [11]:

> The Multilateral Interoperability Program referred to as MIP, is an interoperability organization established by national Command and Control Information Systems (C2IS) developers with a requirement to share relevant Command and Control information in a multinational or coalition environment.

25.6.2 Logistics

One of the big challenges in military operations is to supply troops with all the assets and consumables required in the area of operations. The delivery of goods just in time is a logistical problem. Goods are stored and produced in distributed locations, and they are needed in the operational area, which may be a rather large area. In contrast to non-military logistics, any vehicle or group of vehicles may be subject to threats from explosive devices or ambushes. For this reason, and in addition to civil routing services, chokepoints need to be considered. Convoys can only move at a limited speed. Any stop would subject the convoy to additional danger. Therefore, conflicts at road crossings or along main routes have to be avoided. In addition to ordinary car navigation systems, military logistics require routing with conflict detection. Therefore, logistic tools for convoy planning need to be connected to a network with the functionality to detect conflicts along planned movement routes. This information is of course highly sensitive, because knowledge of the points at which troops are vulnerable could be taken advantage of by hostile forces. This leads to time-dependent geospatial functionality.

The maximal load capacity for tracked and wheeled vehicles and the horizontal and vertical clearances of roads, bridges, and tunnels are restricting factors for military movements. The condition of the road or road structure and possible bypass possibilities may also be taken into account. To determine possible chokepoints and possible hiding places for explosive devices or snipers, the surrounding topography along a planned route has to be investigated. This is done on the basis of high-resolution imagery or high-resolution elevation data or by carrying out infield reconnaissance missions, potentially supported by UAS (unmanned aircraft system) results. Any findings will also improve the utility of the geospatial data for other purposes. This is why geospatial data need to be centralized—the data are then available to other users and other functional area services.

Military convoys are just one aspect of defense logistics. Transportation channels are usually multimodal. Detailed planning and optimization of the various stations used for airlifting, shipping, and overland or helicopter transportation along with the required cargo handling at any change of mode are needed. The tracing of goods is another service where geospatial functionality is involved.

25.6.3 Intelligence, Surveillance, and Reconnaissance (ISR)

Intelligence in the military sense (INTEL) is about gathering, fusing, and analyzing active data. Active data is defined as data that have current relevance and are checked from preferably more than one source. Some data—especially when there are terrorist threats—are related to individuals, while others are related to locations; nonetheless, intelligence information has a spatial property, e. g., the movement profiles of terrorists or incidents with improvised explosive devices (IEDs). Surveillance is the monitoring of behavior, activities, or other changing information. In this respect, surveillance can be seen as the information-capturing component of INTEL. Any sensor that can capture geospatial data can be seen as a surveillance sensor. Since up-to-date data is one of the most important aspects of INTEL, special sensors have been developed for surveillance. Such sensors mostly operate in the optical electromagnetic spectrum and are mounted on unmanned aerial vehicles (drones) or special satellites with a very small ground sample distance and a very agile platform to achieve high repetition frequencies. Military reconnaissance is the act of seeking to determine an enemy's intentions by collecting and gathering information about its composition and capabilities along with pertinent environmental conditions via direct observation. The use of high-resolution imagery for intelligence purposes is also known by the acronyms GEOINT (for *geospatial intelligence*) or IMINT (for *imagery intelligence*) [1].

ISR provides sensors to capture data with geospatial properties. The georeferencing of imagery data can be especially challenging, since ground control points and reliable coordinates in the observed area may be sparse. The exploitation of ISR imagery leads to accurate and current features that are usually used as an overlay for the base data in the common operational or recognized environmental picture.

Military Earth observation satellites, which are also called reconnaissance satellites but are more popularly known as spy satellites, provide intelligence information on the military activities of foreign countries. Enemy missile launches can be detected with so-called early-warning satellites. Nuclear-explosion detection satellites are designed to detect and identify nuclear explosions from space. Photosurveillance satellites provide high-resolution imagery of hostile military activities, e. g., the deployment of intercontinental ballistic missiles. Satellites with sensors in the microwave spectrum provide images even when there is cloud cover or in the dark. Electronic reconnaissance satellites receive and record radio and radar signals while passing over an area of interest. A previous restriction imposed by the US government on the distribution of satellite imagery at a ground sampling distance (GSD) of 0.50 m or less, as defined in the International Traffic in Arms Regulations (ITAR), has since been lifted. However, according to the US Departments of Defense and State, some restrictions for satellite components still apply [12].

Since satellites have limitations in their ability to rapidly deliver imagery of certain areas, and since there are physical limitations on the resolution of imagery data due to the altitudes of satellites, other sensors are required to capture imagery in the battle space. Airborne photogrammetry is a way to capture accurate imagery data at very high resolution. Since it is risky to send pilots over areas in which there are armed conflicts, remotely controlled planes are used to provide reconnaissance information. Such an unmanned airplane or helicopter is called a drone, an unmanned aircraft system (UAS), or an unmanned aerial vehicle (UAV). Different kinds of sensors can be mounted on UAVs. Most UAVs are designed to return and land safely. They either broadcast the gathered data or store it on onboard to be exploited after landing. Precisely orienting the images and orthorectifying them is a challenge, since ground control points are usually not available.

25.6.4 Air-Traffic Control

Air-traffic control (ATC) is a service in which aircraft are directed on the ground and in the air by controllers in control towers. The primary purposes of the ATC system are to prevent collisions between aircraft operating in the system, to organize and expedite the flow of traffic, to provide support for national security, and to assist an aircraft when an emergency has been declared [13].

Preventing collisions is referred to as *separation*; aircraft are prevented from coming too close to each other by imposing minimum lateral, vertical, and longitudinal distances. If the protected airspaces of two aircraft do not overlap, the applicable lateral separation is ensured.

In many countries, ATC services are provided throughout the majority of the country's airspace, and its services are available to all users (private, military, and commercial). When controllers are responsible for separating some or all aircraft, the airspace involved is called *controlled airspace*, in contrast to *uncontrolled airspace*, where aircraft may fly without using an air-traffic control system. Since military and civilian aircraft share the same airspace, the air-traffic control system needs to be well coordinated.

In addition to air-traffic control, air operations need to be planned and coordinated in a command and control system. Just as for ground operations, the COP and REP are therefore required to generate air task orders (ATOs) and air tasking messages (ATMs). The three-dimensional elevation model is important for low-level flight operations, together with a database of flight obstacles. It is intended to integrate air-traffic control, surveillance, air mission control, airspace management, and force management functions into one system.

25.6.5 AWACS

The airborne warning and control system (AWACS) is an aircraft-mounted system that carries out airborne surveillance and command, control, and communication (C3) functions for both tactical and air-defense forces when air-traffic control towers are not available or do not cover the whole area of operation. With a radome 9.1 m (30 ft) in diameter, the AWACS look-down radar has a 360° view of the horizon and a range of more than 320 km at operating altitudes. The radar can detect and track air and sea targets simultaneously. In a tactical role, AWACS can detect and track hostile aircraft operating at low altitudes over any terrain, and can identify and control (monitor) friendly aircraft in the same airspace. When used in a strategic defense role, AWACS provides the means to detect, identify, track, and intercept airborne threats [14].

The radome scans at six revolutions per minute. The radar is multimode and uses interleaving and deinterleaving algorithms. The main operating modes are pulsed Doppler nonelevation scan (PDNES) for the surveillance of airborne targets, pulsed Doppler elevation scan (PDES) to determine the target elevation, beyond-the-horizon pulse radar mode, receive-only mode for passive operation, maritime mode using a very short pulse width for the detection of surface ships, and standby mode. AWACS provides a current map of objects identifiable by radar.

25.6.6 Confidence-Building Measures and Disarmament

Adherence to arms control is the best way of achieving mutual confidence in the peaceful intentions of other states. Disarmament treaties need to be observed to avoid power imbalances. The Organization for Security and Co-operation in Europe (OSCE) and Open Skies are examples of arms control operations. These peacetime operations produce, partly as byproducts, a huge amount of geospatial data and imagery.

The OSCE is a security-oriented intergovernmental organization under the United Nations Charter. There are currently 57 states in Europe, the Caucasus, Central Asia, and North America that participate in the OSCE. Its mandate includes issues such as arms control, human rights, freedom of the press, and fair elections. The end of the Cold War resulted in a huge amount of surplus weapons becoming available illegally. The OSCE helps to stop the spread of such weapons and offers assistance with their destruction. With its expertise in conflict prevention, crisis management, and early warning, the OSCE contributes to worldwide efforts to combat terrorism. It works to prevent conflicts from arising and to facilitate lasting comprehensive political settlements for existing conflicts. It also helps with the process of rehabilitation in post-conflict areas. The OSCE's Forum for Security Cooperation provides a framework for political dialog on military reform, while practical activities are conducted by field operations as well as the Conflict Prevention Centre. OSCE police operations are an integral part of the organization's efforts towards conflict prevention and post-conflict rehabilitation [15].

The Treaty on Open Skies establishes a regime of unarmed aerial observation flights over the territories of its signatories. Thirty-four state parties have ratified the treaty so far. It is designed to enhance mutual understanding and confidence by giving all participants, regardless of size, a direct role in gathering information through aerial imaging on military forces and activities of concern to them. Open Skies is one of the most wide-ranging international arms-control efforts to promote openness and transparency in military forces and activities to date [16].

Open Skies aircraft may have video, optical panoramic, and framing cameras for daylight photography, infrared line scanners to provide a day/night capability, and synthetic-aperture radar (SAR) to provide a day/night all-weather capability. Photographic image quality permits the recognition of major military equipment (e. g., it permits a state party to distinguish between a tank and a truck), thus allowing significant transparency of military forces and activities. Sensor categories may be added and capabilities improved by agreement among state parties. All equipment used in Open Skies must be commercially available to all participants in the treaty.

Each state party is obligated to receive observation flights per its *passive quota* allocation. Each state party may conduct as many observation flights—its *active quota*—as its passive quota. The Russian Federation and the USA each have an annual passive quota of 42, and other state parties have quotas of between 2 and 12. Each October, the parties negotiate the distribution of active quotas for the following calendar year. Over 100 observation flights are conducted each year.

Imagery collected from Open Skies missions is available upon request to any state party for the cost of reproduction. As a result, the amount of data available to each state party under the treaty quota system is much greater than the amount each party can collect itself.

25.6.7 Electronic Warfare Analysis Systems

Electronic warfare (EW) refers to any action involving the use of electromagnetic waves to control the spectrum or to prevent other users from using electromagnetic waves for communication or navigation by jamming certain frequencies of the electromagnetic spectrum. EW analysis systems provide the capability to analyze electromagnetic radiation on the battlefield (communication equipment and radar) and determine the positions and types of radiation sources present. EW analysis systems use geospatial analysis functionality to calculate the theoretical coverage of friendly and enemy radars, radios, passive sensors, or jammers.

The main factors that influence the propagation of electromagnetic radiation, depending on its wavelength, are (in order): the terrain, weather, and ground coverage. Therefore, EW analysis systems require digital elevation, meteorological, and topographic data. The resolution of the required data depends on the wavelength of the electromagnetic radiation under investigation. Very-high-frequency/ultrahigh-frequency (VHF/UHF) waves propagate in a completely different way to microwaves such as those used for radar and Global Navigation Satellite Systems (GNSS).

A key function of the system is the ability to calculate electromagnetic lines of sight (LOS) for radar and radio at different frequency bands using propagation models that take into account atmospheric effects. Terrain masking is a general term for the process of determining what is visible or detectable and what is not visible or detectable in a given area of terrain from a given location. Its primary use is in the evaluation of terrain for natural and tactical obstructions to support LOS intervisibility, threat detection, and sensor, weapon, or communications emplacement. The first use deals specifically with the optical (visible light) spectrum while the latter two consider the electromagnetic (EM) spectrum. The differences between optical and EM calculations are substantial. Light-sensitive sensors such as those in cameras can differ from sensors capable of detecting non-visible radiation in terms of what they can see. The latter require a set of complex algorithms and a database of parameters for individual sensors and communication equipment. Refractions and reflections may also be taken into consideration.

Wave propagation algorithms have to interact with geospatial data because propagation is dependent on spatial and material properties along the radiation path, e. g., the spatial distribution of atmospheric conditions. Visualization of wave propagation results needs to be provided on top of the geospatial base data or in combination with other operational or tactical features for further evaluation or mission planning.

25.6.8 Embedded Systems

The most obvious defense application of geospatial data and the geospatial analysis functionality described in the previous sections is in command and control systems. Another widely but not yet fully exploited area of GIS utilization for defense is in embedding data and functionality into assets, vehicles, planes, or weapon systems.

Television (TV) imaging (electro-optical imaging) is a navigation system for cruise missiles in which an electro-optical seeker scans a designated area for targets via optical imaging. Once a target is acquired, the missile locks onto it. TV imaging does not depend on the target's heat signature and thus can be used against low-heat targets. The system, however, has the drawback that the target must be *seen* by the missile, which limits its range of action [17].

25.6.9 Training and Simulation

Soldiers need to be trained in the use of their equipment. Cooperation between different branches and units must also be practiced. However, military exercises are very expensive, time-consuming, and possibly hazardous, and not everything can be exercised in field maneuvers. Virtualization and simulation can be less expensive and safer substitutes for real exercises. Simulations can be conducted with artificial geospatial data or with real operational data for the anticipated operational area. With 3-D functionality and physical models of weapon systems, simulations can be conducted in virtual reality. Flight simulation is not only big business for the gaming industry; it is also very important for teaching or instructing pilots to use new aircraft, tanks, vessels, helicopters, and other platforms. Simulation has the advantage that training can be recorded in detail and analyzed later to gain knowledge. Also, interoperability between different systems can be tested under real deployment conditions.

With the increasing importance of network-centric warfare, virtualization of training and simulation is also becoming more important.

25.7 Military Standards

25.7.1 WGS84

National Geospatial-Intelligence Agency (NGA) Technical Report TR8350.2 defines World Geodetic System 1984 (WGS84) [18]. This definition of a global geodetic datum is used for the Global Navigation Satellite Systems (GNSS). It provides the base reference frame for most military maps and is used in conjunction with a coordinate system to create the coordinate reference systems (CRSs) used for most geospatial datasets. Maps and geodata produced by defense mapping agencies at the national level may be based on national CRSs that are different from WGS84. For coherence with other military data, these maps and geodata must be transformed to a uniform CRS. Military geodata are provided in coordinates based on WGS84 to allow direct data usage without any further loss of positional accuracy.

25.7.2 Coordinate Systems

Different map projections are defined for the spheroid defined by WGS84, and these projections are commonly used for military applications. A list of the most important coordinate systems is given in Table 25.3.

The Universal Transverse Mercator (UTM) and Universal Polar Stereographic (UPS) map projections as well as the Military Grid Reference System (MGRS) are military standards defined by the National Geospatial-Intelligence Agency (NGA) Technical Manual 8358.1 [19] (Chap. 8).

25.7.3 DIGEST

Due to the lack of an open, vendor-independent standard to exchange digital geospatial data, a group of international defense mapping organizations founded the Digital Geographic Information Working Group (DGIWG) in 1983. The Digital Geographic Information Exchange Standard (DIGEST) was their first specification. DIGEST went through several edition cycles, and its latest version (edition 2.1) was made publicly available in the year 2000. Meanwhile, the DGIWG changed its name to the Defense Geospatial Information Working Group and is working on a suite of modern standards using the approach employed by ISO/TC 211 and the OGC (Chap. 13). The DGIWG is preparing to retire DIGEST. Part 4, the FACC (Feature Attribute Coding Cat-

Table 25.3 Coordinate systems typically used in military GIS applications

Geographic coordinates	For nonprojected coordinates, the latitude and longitude geographic coordinate system is used
UTM	The UTM coordinate system divides the world into 60 zones extending from 80° S to 84° N in latitude, each of which is 6° in longitude wide. The polar regions are excluded. The first zone starts at the International Date Line (longitude 180° with respect to the Greenwich Central Meridian) and the zones proceed eastward
UPS	The UPS coordinate system defines a conformal projection for the polar regions. It extends to 83°30′ N and 79°30′ S to provide a 30′ overlap with the UTM
MGRS	MGRS or UTMREF provides a compact labeling convention for UTM or UPS coordinates. It combines the grid zone designator with a two-character identifier for 100 km squares and the rectangular coordinates within a square

alogue), has already been sunsetted and replaced by DFDD, the DGIWG Feature Data Dictionary (DFDD).

DIGEST (edition 2.1) is also a standard (named STANAG 7074) within NATO. Although DIGEST is deprecated, it gives a good insight in the needs to structure geospatial data for military purposes. DIGEST is divided into four parts [20]:

- Part 1 contains a general description of the standard.
- Part 2 provides the theoretical model and an exchange structure, with four different encapsulation specifications provided in annexes. The encapsulation utilizes international standards for encoding.
- Part 3 defines basic data types, including coordinate representation and geodetic codes and parameters.
- Part 4 is the Feature Attribute Coding Catalogue (FACC). This data dictionary was developed as a comprehensive coding scheme for feature types, feature attributes, and attribute values. It has since been replaced by the DFDD, a subset of the DGIWG's Defense Geospatial Feature Concept Dictionary (DGFCD; see Sect. 25.8.3).

Vector data in DIGEST are organized according to their topological relationships. Features may have a complex structure, i.e. features may be composed of parts that are geospatial features on their own. Additionally, depending on the application complexity four levels of topological structures are supported (Fig. 25.4):

- Spaghetti vector data (level 0 topology)
- Chain-node vector data (level 1 topology)
- Planar graph vector data (level 2 topology)
- Full topological vector data (level 3 topology).

The features can either be complex (i. e., they are composed of subordinate features) or they are simple. *Simple* in DIGEST means that they have neither superordinate nor subordinate features. In contrast to the OGC's Simple Feature Specification (see Chaps. 3, 13, and 30), *simple* in this context is not with respect to it's geometry. The features are related to their geometry, and the geometric primitives are then related to the topological primitives.

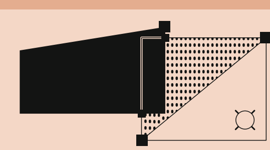
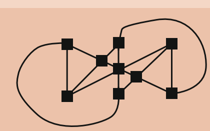
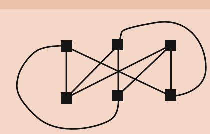
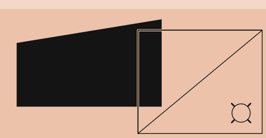
Level	Name	Primitives	Description	Example
3	Full topology	Connected nodes, entity nodes, edges, and faces	The surface is partitioned by a set of mutually exclusive and collectively exhaustive faces. Edges meet only at nodes	
2	Planar graph	Entity nodes, connected nodes, and edges	A set of edges and nodes where, when projected onto a planar surface, the edges meet only at nodes	
1	Nonplanar graph	Entity nodes, connected nodes, and edges	A set of entity nodes and edges that may meet at nodes	
0	Boundary representation (spaghetti)	Entity nodes and edges	A set of entity nodes and edges. Edges contain only coordinates	

Fig. 25.4 Levels of topology in DIGEST [21]

For a long time, Part 4 of DIGEST—the FACC—was the dictionary for application schemas. It provided a list of feature type names together with a code of five characters and a definition (e. g., BH140 for feature type *river/stream*, which was defined as *a natural flowing watercourse*). Additionally, FACC provided attribute names and possible attribute values (e. g., attribute code MCC for *material composition category*, which had a long list of possible values such as *marble* with code 60). The attributes and attribute values were not feature specific and were meant to be used to develop a feature catalog or application schema by assigning the correct attributes to the right features.

The FACC was replaced by the DGIWG Feature Data Dictionary (DFDD) (Sect. 25.9.1).

25.8 New Generation of Military Standards

The standards used in DIGEST do not fulfill the requirements of a flexible network-centric architecture for data modeling and geospatial data exchange. The DGIWG therefore decided to develop a new suite of standards based on ISO standards of the 19100 series and on OGC standards. In order to be compliant with commercial off-the-shelf products, the new standards will mainly be profiles (i. e., distinct subsets or subsets with potential extensions) of the base standards. ISO standards are endorsed or adapted to fulfill military requirements. For example, several profiles have been developed for ISO 19136 *Geography Markup Language (GML)* to reflect the requirements of defined military use cases. The DGIWG GML profiles also provide profiles that reflect the DIGEST way of expressing explicit topology.

In addition to the standards for exchanging geospatial data, the new generation of DGIWG standards also support (web) services. This is a requirement from the military community to enable geospatial information services in network-centric warfare.

25.8.1 DFDD

A new data dictionary was developed as a replacement for the FACC and is maintained as a register, i. e., users can trace back any changes to the register. The new feature concept dictionary is called the DFDD. The DFDD is maintained in a register that is accessible online [22]. Because deprecated items get retired rather than deleted, older versions of the register can still be accessed.

The DFDD does not require the usage of the feature and attribute codes of the FACC anymore. For backwards compatibility, FACC codes are maintained in the register. Applications are supposed to use the meaningful versions of feature type names, e. g., *River* instead of *BH140*. The list of feature and attribute types has been revised to remove ambiguities. New items have been added since the transition from the FACC.

25.8.2 Registers

Registers other than the DFDD have been established and are publicly available on the DGIWG website [23]. The registers with high priority are those for:

- Terminology
- Feature and attribute data
- Geodetic codes and parameters
- Metadata elements
- Portrayal symbols and portrayal rules.

Other registers (e. g., for data quality measures or product specifications) may follow.

The registers adhere to ISO 19135, the standard for item registration procedures (Chap. 16).

25.8.3 NGIF/DGIF

As in the scenarios described in Sect. 25.2, nations capture their own geospatial data based on their national requirements and their capabilities. In order to do so they define and implement national conceptual schemas. These schemas are often formulated using feature types, attribute names, or attribute codes in their national language. In common operations, to support the principle of operating off the same map, the geospatial data needs to be semantically translated and transformed. If we assume that n nations are cooperating, and that data exchange occurs in both directions, we find that $n(n-1)$ semantic translators are needed, which increases quadratically with the number of conceptual models involved (Fig. 25.5). Despite the effort, this leads to tremendous semantic interoperability issues. To avoid these issues, the DGIWG developed a comprehensive conceptual model for data exchange that reduces the number of semantic translators required to $2n$ (Fig. 25.6). All specifications that address the topic of semantic interoperability together comprise the Defense Geospatial Information Framework (DGIF). Its NATO counterpart, which may not adopt all the concepts of the DGIF, is called the NATA Geospatial Information Framework (NGIF) [24].

The DGIF comprises three specifications, called artefacts, that together lead to Data Product Specifications (DPSs) in

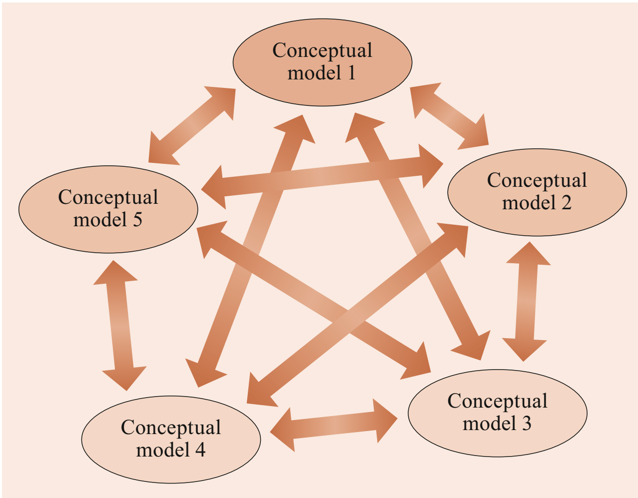


Fig. 25.5 Requirement for $n(n-1)$ semantic translators

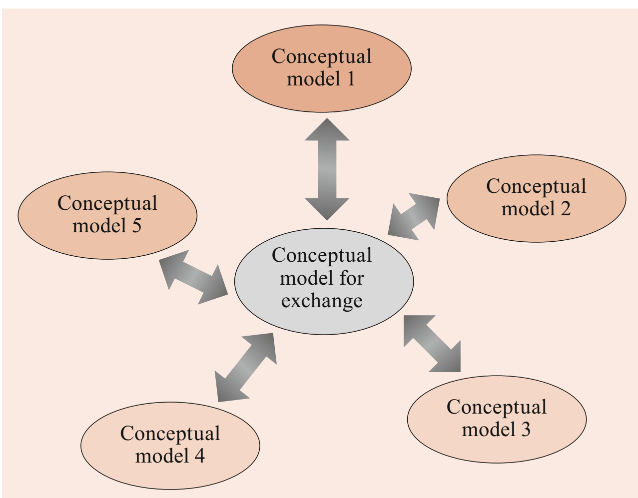


Fig. 25.6 Reducing the requirement to $2n$ semantic translators by defining a common semantic exchange model

compliance with ISO 19131:2007. The three specifications are the DGFCD, DGIM, and DGRWI (see explanations below).

Defense Geospatial Feature Concept Dictionary (DGFCD)

Each feature type, attribute name, attribute value, data type, and unit of measurement to be used to define a conceptual model is defined in the DGFCD with an alpha code, their name, a definition, a description, and potential alias. This comprehensive list represents the concepts of many user communities as it incorporates various established dictionaries, such as the Aeronautical Information Exchange Model (AIXM), International Hydrographic Organization Standard No. 100 (IHO S-100), and NATO Additional Military Layers (AML). It builds the vocabulary needed to set up the Defense Geospatial Information Model and lays out a common understanding of the concepts behind names or codes—a prerequisite for semantic interoperability.

Defense Geospatial Information Model (DGIM)

How the different concepts relate to each other needs to be defined in a conceptual model. The DGIF uses Unified Modeling Language (UML) as a graphical notation to define these relations. As UML supports the object-oriented modeling paradigm, the class diagram provides the most important notation to define the conceptual model of the DGIM. Feature types are represented as classes and attributes are class properties that belong either to an enumeration with defined attribute values provided by the DGFCD or have a certain datatype, also provided by the DGFCD.

Due to its complexity, the DGIM is not intended to be implemented in its entirety as a logical model for a geodatabase. This conceptual model is designed to serve as a conceptual model for exchange (Fig. 25.6). Defined subsets, so-called profiles, may be defined for certain user communities. In that respect, rules for mapping from a national or proprietary conceptual model to the DGIM and vice versa need to be defined. Following the model-driven approach, it is possible to derive exchange formats from the DGIM or any profile, e. g., to define a Geography Markup Language (GML) application schema.

Defense Geospatial Real World Object Index (DGRWI)

A conceptual schema depends heavily on the viewpoint of the user community regarding the universe of discourse or the intended application. In that respect, it cannot be unique or perfect or comprehensive. The viewpoint determines, e. g., what the user community considers a feature type, as this is also a matter of granularity or level of detail. To identify how a concept is represented in the DGIM, a mechanism to search for alternative concepts is required. The DGRWI connects colloquial expressions for concepts with their counterpart in the DGFCD. For example, the word *church* has different meanings depending on the context and therefore does not represent a feature type in the DGFCD. As the DGFCD is based on the physical appearances of geographical phenomena, *church* would be represented as a building with the property religiousFacilityType—a datatype enumeration with a list of possible attribute values, such as church or chapel. In order to distinguish the religion that this religious facility belongs to, the property religiousDesignation—which lists the possible religions—is provided.

25.9 Military Datasets

A selection of standardized geospatial data specifications are introduced in this section.

25.9.1 Level of Detail and Resolution

For operations at strategic, operational, and tactical levels, continuous homogeneous geospatial data with various resolutions are required. For this reason, various series of geospatial data with different levels of detail have been specified and captured (Military Committee MC 296/2: NATO Geospatial Policy [25]). These data reflect the typical map scales that are well established in defense mapping:

- Level 0—1 : 500,000 and smaller
- Level 1—1 : 250,000
- Level 2—1 : 50,000
- Level 3—highest resolution.

In addition to the data specifications, geospatial products may be defined for each level. Geospatial products consist of a specified way of portraying the geospatial data and preassigned formats, possibly including the standardized symbology. The assumption behind geospatial products is that all users perceive the data in the same way. This avoids arbitrary representations of the same data, thus preventing confusion. As with printed maps, which have well-established legends, users become familiar with the appearance of geospatial data, avoiding arbitrariness.

25.9.2 Paper Maps and Their Scanned Georeferenced Counterparts

The easiest way to reuse existing paper maps in geospatial information systems is by scanning and georeferencing the map and cropping the marginalia from the actual map content. When performed in connection with smart data management, including tiling and spatial indices, this leads to a seamless dataset for background display.

Well-established map products that are still in use and printed for field operations are now available as raster products. A selection of the huge range of map products available can be found in Table 25.4. Special products for specific applications are not listed, and only the major nautical navigation charts are listed.

Table 25.4 A selection of the paper map series used in defense applications

TLM	Topographic Line Map	1 : 50,000
JOG	Joint Operations Graphics	1 : 250,000
TPC	Tactical Pilotage Chart	1 : 500,000
ONC	Operational Navigation Chart	1 : 1,000,000
JNC	Jet Navigation Chart	1 : 2,000,000

25.9.3 Satellite Imagery and Aerial Photographs

Geospatial imagery is the most requested source of information for defense applications. Imagery from various sensors can be used, depending on requirements. Radar sensors for imagery as well as for elevation models derived from radar phase measurements have the advantage of being active and are therefore not dependent on the presence of daylight. Also, due to the ability of microwaves to penetrate clouds, they are also excellent for areas and seasons with a high likelihood or percentage of cloud coverage. However, because they transmit radar signals when active, they can easily be detected and located by other sensors and consequently destroyed.

Passive sensors in the visual part of the electro-magnetic spectrum provide either still imagery or videos. They can be panchromatic, they can detect red–green–blue (RGB) excluding or including infrared, or they can be multispectral. The raw imagery provided by the sensor needs to be orthorectified, i. e., georeferenced and corrected for displacements due to height differences originating from the central perspective. Whereas calibrated sensor parameters are usually provided with images or videos, it may be very challenging to identify control points (if they exist at all) and to apply a sufficiently accurate elevation model. Imagery sensors may be mounted along with a light detection and ranging (LIDAR) laser ranging system that can solve the elevation issue as long as the LIDAR sensor orientation is known to sufficient accuracy.

Sensors for radar, optical imagery, and LIDAR can be mounted on different platforms: either satellites or manned or unmanned aerial vehicles (Sect. 25.6.3).

Commercial satellite imagery is used to provide spatially inclusive and comprehensive image coverage of the area of interest. It can then be refined with more accurate and current imagery, possibly with a higher resolution.

Different imagery formats are applied. Due to the huge amount of data, high compression rates and fast image reconstruction algorithms are an advantage. For easier handling, embedded georeference data and metadata are preferred over separate files.

25.9.4 Vector Map

The Vector Map (VMAP) product is a collection of vector-based geospatial data at low, medium, and high resolution. These data are separated into nine thematic layers and are topologically structured. A VMAP is sometimes also referred to as a vector smart map.

Features belong to one of nine thematic layers: boundaries, population, industry, transportation, utilities, hydrography, vegetation, physiography, and elevation. A data quality index is also provided. Feature types and attribution conform to the deprecated Feature Attribute Coding Catalogue (FACC). VMAP data are encoded in the VPF (Vector Product Format) structure as specified by Annex C of DIGEST 2.1. Feature coordinates are stored in geographic coordinates with decimal degrees based on WGS84. Height information is given in meters and is referenced to a vertical datum: mean sea level (MSL).

VMAPs are produced at different levels of detail and resolution. Only VMAP level 0 and level 1 are intended to have worldwide coverage. VMAP level 2 and level 3 are only produced for designated areas where a requirement is anticipated.

VMAP level 0 is an updated and improved version of the Digital Chart of the World (DCW). VMAP level 0 and VMAP level 1 can be downloaded for free from various websites. VMAP level 1 data represent the content of the Joint Operations Graphics (JOG) at a scale of 1 : 250,000. Level 1 is divided into an irregular mosaic of 234 geographic areas such that each area fits on a single CD-ROM. Only 57 of these areas are currently available for public download.

VMAP level 2 corresponds to the content of a Topographic Line Map (TLM) at a scale of 1 : 25,000. VMAP level 3 is called the Urban Vector Map (UVMAP). It contains high-resolution two-dimensional feature and attribute content from city graphics (town plans).

VMAP level 2 will be substituted by the Multinational Geospatial Coproduction Program (MGCP) (see later), with anticipated worldwide coverage.

25.9.5 Digital Terrain Elevation Data

Digital Terrain Elevation Data (DTED) provides heights above sea level across a regular grid (matrix). It is produced at three different levels of detail. The three classes of DTED are known as DTED level 0, DTED level 1, and DTED level 2 [26].

The horizontal datum for DTED data is the World Geodetic System (WGS84). The vertical datum is mean sea level as determined by Earth Gravitational Model 1996 (EGM96).

Due to the convergence of meridians towards the poles, the matrix interval depends on latitude. Tables 25.5–25.7 show the latitude dependence of the matrix interval for the three levels [27]. DTED level 3 is reserved for very-high-resolution elevation data. Since 1″ is approximately 30 m in the north–south direction, the three levels are no longer sufficient for the requirements of modern C2 systems.

Table 25.5 Matrix intervals for DTED level 0

Zone	Latitude	Matrix interval		
		Latitude		Longitude
I	0–50° north–south	30″	×	30″
II	50–70° north–south	30″	×	60″
III	70–75° north–south	30″	×	90″
IV	75–80° north–south	30″	×	120″
V	80–90° north–south	30″	×	180″

Table 25.6 Matrix intervals for DTED level 1

Zone	Latitude	Matrix interval		
		Latitude		Longitude
I	0–50° north–south	3″	×	3″
II	50–70° north–south	3″	×	6″
III	70–75° north–south	3″	×	9″
IV	75–80° north–south	3″	×	12″
V	80–90° north–south	3″	×	18″

Table 25.7 Matrix intervals for DTED level 2

Zone	Latitude	Matrix interval		
		Latitude		Longitude
I	0–50° north–south	1″	×	1″
II	50–70° north–south	1″	×	2″
III	70–75° north–south	1″	×	3″
IV	75–80° north–south	1″	×	4″
V	80–90° north–south	1″	×	6″

25.9.6 Multinational Geospatial Coproduction Program

The Multinational Geospatial Coproduction Program (MGCP) comprises a coalition of nations participating in the capture of high-resolution geospatial vector data for all land masses. The MGCP application schema is based on the DFDD (Sect. 25.8.1). The captured MGCP data are uploaded to the International Geospatial Warehouse (IGW) for storage and exchange with other coproducers. The IGW was established and is maintained by the US National Geospatial-Intelligence Agency (NGA). Quality control is conducted at the IGW so that only products adhering to the agreed data quality acceptance level will be integrated into the database [28].

The members of the MGCP consist of 31 countries, including 11 *lead nations* and several industry partners who provide technical assistance. The MGCP‘s goal is to work globally at the 1 : 50,000 or 1 : 100,000 (for rural or deserted areas) scale. Production is divided into cells that are 1° square [29].

The IGW is a secure website that allows access only to authorized registered users. Any uploaded cell must be provided with appropriate metadata. Metadata must also be provided for subregions of the cell if different data sources are used.

The exchange of data is based on a credit/debit system. This means that the amount of geospatial data that can be extracted by a nation from the portal depends on the amount of cells it has provided. Some Caribbean cells have been made publicly available for free download to support humanitarian relief after a major earthquake and several hurricanes [30].

25.9.7 Automated Air Facilities Intelligence Files

Automated Air Facilities Intelligence Files (AAFIFs) describe the physical characteristics of airfields. They contain evaluated information in eight categories and 65 subcategories, and more than 400 required attributes. Elements range from a single-character-coded field to textual fields of several hundred characters. The categories consist of aircraft movement surfaces (runways, taxiways, aprons, etc.), facilities, support equipment, services, operations, navigational aids/communications, transportation, and other items for approximately 48,000 airfields worldwide [31].

Feature coordinates are given as latitudes and longitudes based on WGS84, with the mean sea level as the vertical datum.

25.9.8 Digital Vertical Obstruction File

The Digital Vertical Obstruction File (DVOF) is managed as a web service for storing, retrieving, and distributing data concerning known vertical flight obstructions worldwide. Vertical obstructions are manmade features on the Earth's surface that are sufficiently tall to pose a potential hazard to flight. Examples of such features include antennas, buildings, pylons, smokestacks, storage tanks, towers, and power lines. A vertical obstruction feature may be represented geometrically as either a point, line, or area [32].

This information can be retrieved by giving a center point and radius, a minimum bounding rectangle, a polygon, or a flight corridor and by filtering on various parameters such as height, obstruction type, and country. The data can then be downloaded in various formats to be used in a variety of platforms and systems. DVOF data are used in many different applications, ranging from building navigation charts to designing terminal approach procedures.

25.9.9 Digital Nautical Chart

The Digital Nautical Chart (DNC) is an unclassified, vector-based digital database containing significant maritime features essential for safe marine navigation. It is produced by the National Geospatial Intelligence Agency. The initial data collected in the database derive from a portfolio of approximately 5000 nautical charts, and will ultimately provide global marine navigation between the latitudes 84° N and 81° S and support a variety of geographic information system applications [33].

The DNC is produced in the standard Vector Product Format (VPF). It contains four categories based on the scale and purpose of the source charts, including harbor, approach, coastal, and general charts (ranging from the largest to the smallest scale, respectively).

The features depicted are organized thematically into 12 layers: cultural landmarks, earth cover, environment, hydrography, inland waterways, land cover, limits, aids to navigation, obstructions, port facilities, relief (bathymetry), and data quality.

25.9.10 Tactical Ocean Data

Tactical Ocean Data (TOD) provides a series of vector-based digital products for naval surface and subsurface navigation. The series comprises several levels. TOD level 0 (TOD0) represents charts that portray naval operating areas (OPAREAs) and naval exercise areas (NAVEX) in a format suitable for computerized navigation. TOD0 is designed to be used in conjunction with the DNC for complete navigation information. TOD0 also functions as a general-purpose global database designed to support geographic information system (GIS) applications. Other levels provide sea-bottom contour charts (TOD1), bathymetric navigation planning charts (TOD2), the seafloor configuration in shallow-water areas (TOD3), and very detailed bathymetric data for submarine hull integrity test sites (TOD4). The TOD specification provides a description of the content, accuracy, data format, and design of the TOD database. In addition, it portrays strategic information to support naval operations. TOD is only available to authorized customers [34].

25.9.11 Tactical Pilotage Chart

A Tactical Pilotage Chart (TPC) is designed to provide an intermediate-scale translation of cultural and terrain features for pilots/navigators flying at medium to very low altitudes (below 500 ft above ground level) or low altitude–high speed operations [35]. The successful execution of a low-altitude mission depends entirely upon the visual and radar identification of ground features as landmarks for flight navigation. For low-altitude flights, the angular velocity of ground features with respect to the aircraft is very high and increases nonlinearly as the aircraft approaches the landmark. Therefore, the pilot or navigator has little time for recognition.

Depth of vision is also restricted because of the increased effect of perspective resulting from the closeness of the aircraft to the terrain. Depth of vision may additionally be shortened by ground fog, haze, and other atmospheric factors. Under these visibility restrictions, the pilot or navigator needs to recognize checkpoint features near the horizon ahead of the aircraft in order to visually identify ground objects rapidly for navigation. To facilitate the recognition of checkpoints at first glance, the pilot or navigator must have a preconceived mental image of each successive checkpoint feature. The pilot must have an appreciation of the design and basic characteristics of these checkpoints and know when (in seconds) and where (relative to the speed of the aircraft) they will be overflown [35].

Therefore, the selection and portrayal of ground features should be based upon the requirement for rapid visual recognition of significant chart details as seen from a low perspective angle. In that respect, TPC symbology differs from topographic or other navigation charts.

25.9.12 Urban Models

There are several use cases where a 2-D map view or a simple elevation model is insufficient. In a 3-D world with anthropogenic and natural objects, it is possible to, e. g., perform line- or area-of-sight analyses with much higher reliability than achievable with just an elevation model. For detailed tactical planning and urban warfare or evacuation operations, urban models are required. The required level of detail for the urban model depends on the use case. Also, a mix of different levels of detail may be possible because not all buildings may be required in full detail. For instance, in an embassy evacuation scenario, interior information may only be required for the embassy building, and the facade and roof construction may be modeled in detail for buildings surrounding the embassy whereas buildings further away may be represented as simple extruded blocks. It is quite obvious that the interiors of buildings in a hostile environment cannot easily be captured. Using high-resolution imagery with high redundancy captured from different angles together with laser scanning data, it is possible to automatically record an urban model without requiring (or requiring only very limited) semantic information.

Urban Vector Map (UVMAP) extends the series of VPF product specifications (Sect. 25.9.4). UVMAP requirements have been stated in the year 2000. Meanwhile, civilian specifications such as CityGML [36] can be used as commercial-off-the-shelf replacement for implementing this military standard. CityGML fulfills the requirement for network enablement.

25.9.13 General Regularly-distributed Information in Binary

General Regularly-distributed Information in Binary (GRIB) is a binary data format, internationally defined by the World Meteorological Organization (WMO), for the exchange of meteorological prediction data [37]. Weather-related data are encoded as gridded coverage. A huge part of this standard is used to define coordinate reference systems and the encoding of the matrix data georeference. Other parts deal with meteorological data dictionaries that define, e. g., different types and amounts of precipitation and cloud types, including the specific codes used in data exchange: 0—clear; 1—cumulonimbus; 2—stratus; 3—stratocumulus; 4—cumulus; 5—altostratus; 6—nimbostratus; 7—altocumulus; 8—cirrostratus; 9—cirrocumulus; 10—cirrus.

Data in GRIB format are used for air navigation, operational planning, and to create the recognized environmental picture. The codes used for GRIB edition 2 are constantly maintained and updated by the WMO. Nevertheless, multidimensional array-oriented scientific data can also be stored and disseminated using the OGC standard NetCDF [38].

Conclusion

Defense has been, and still is, a driving factor for the development of GIS. The demanding requirements for military systems expedite the development of geospatial functionality and network-centric geospatial capabilities. Defense mapping agencies—often called geospatial information or geospatial intelligence agencies nowadays—capture huge amounts of geodata. It is crucial to get the data or information derived from the data to the commander or soldier involved in an operation as fast as possible. Very often, base geodata are not classified and are made available on the Internet for free download—especially to support areas that have suffered natural or other humanitarian disasters.

The heterogeneity and complexity of the different systems in defense applications require lossless exchange of information, which leads to a strong need for interoperability. Directly or indirectly georeferenced data need to be fused and collectively analyzed in order to obtain new information that can provide knowledge superiority during in-theater decision-making. Given the frequent rotation of military personnel in theater, the time required to learn how to use GIS tools needs to be shortened. The presentation and user interface of the map should be unambiguous, standardized, and correspond to the user's experience. Due to the need for interoperability and standard operating procedures, it is no surprise that defense agencies are key contributors to the standardization of GIS. This standardization covers concepts, interfaces, formats, and product specifications, along with all their components. This chapter has provided the fundamentals needed to understand the military requirements of GIS. It has given an overview of the use cases and requirements of GIS in defense and the specifications used by the defense community. The chapter has also introduced the concepts of DIGEST, a retired standard of diminishing importance, and discussed emerging and new concepts that will provide the basis for the utilization of GIS in defense applications in the future.

References

1. Joint Chiefs of Staff used by the United States: Joint Task Force Headquarters, Joint Publication 3-33. Government Printing Office, Washington (2018). https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_33.pdf
2. Defence Geospatial Information Working Group: Operational Scenarios. DI JE, London (2011). available from https://www.dgiwg.org/technical-strategy#Operational_Scenarios
3. UN: United Nations Security Council Resolutions. United Nations, Geneva (2019). https://www.un.org/securitycouncil/content/resolutions-0
4. Mittu, R., Segaria, F.: Common Operational Picture (COP) and Common Tactical Picture (CTP) Management via a Consistent Networked Information Stream (CNIS). Naval Research Laboratory, Washington (2000). available from https://www.nrl.navy.mil/itd/imda/sites/www.nrl.navy.mil.itd.imda/files/pdfs/COP-CTP-CNIS.pdf
5. Department of Defense: MIL-STD-2525D: Joint Military Symbology. Department of Defense Interface Standard, Washington (2008). available from https://www.jcs.mil/Portals/36/Documents/Doctrine/Other_Pubs/ms_2525d.pdf
6. United Nations: Military Symbols Handbook. Department of Peacekeeping Operations, New York (2000). http://www.deadlyriver.com/mmaps/UNmilitarysymbols_handbook_DPKO_2000.pdf
7. Balsinhas, J., Teufert, J.F.: GIS technologies in support of the NATO recognized environmental picture. In: Proc. ESRI Int. User Conf. ESRI, Redlands (2006). available from http://proceedings.esri.com/library/userconf/proc06/papers/abstracts/a2401.html
8. Teufert, J.F., Trabelsi, M.: Development of NATO's recognized environmental picture. Proc. SPIE **6201**, 62011D (2006)
9. NATO NEWS: NATO BC-AIS Core OGC Newsletter, October. Open Geospatial Consortium, Wayland (2005). available from http://www.opengeospatial.org/pressroom/newsletters/200510/#C4
10. Stenbit, J.P., Wells II, L., Alberts, D.S.: NATO Code of Best Practice for C2 Assessment. OASD, CCRP, Washington (2002). available from http://dodccrp.org/files/NATO_COBP.pdf
11. Europe, N.A.T.O.: Multilateral Interoperability Programme. NATO, Brussels (2005). available from https://public.mip-interop.org/sites/mip/Pages/Home.aspx
12. Departments of Defense and State: Final Report to Congress (1248 report) (2012). https://www.globalsecurity.org/space/library/report/2012/1248-space-exports.pdf
13. USDOT: Air Traffic Control. US Department of Transportation, Washington (2010). available from http://tfmlearning.faa.gov/publications/atpubs/ATC/atc0201.html
14. Net Resources International: E-3 AWACS Airborne Warning and Control System (2010). http://www.airforce-technology.com/projects/e3awacs/
15. OSCE Secretariat Vienna: Organization for Security and Cooperation in Europe (2010). http://www.osce.org/
16. United Nations Office for Disarmament Affairs: Treaty on Open Skies (Entered into force: 1 January 2002) (2002). https://treaties.unoda.org/t/open_skies
17. Nebylov, A.V., Watson, J. (eds.): Aerospace Navigation Systems. John Wiley & Sons, Chichester (2016)
18. Department of Defense: World Geodetic System (1984). http://earth-info.nga.mil/index.php?dir=wgs84
19. DMA: Datums, Ellipsoids, Grids, and Grid Reference Systems, DMA Tech. Manual 8358.1. Defense Mapping Agency, Fairfax (1990). available from http://earth-info.nga.mil/GandG/publications/tm8358.1/toc.html
20. North Atlantic Treaty Organization (NATO): STANAG 7074 Digital Geographic Information Exchange Standard (Digest) (1998). https://standards.globalspec.com/std/233133/STANAG%207074
21. DMA: Military Standard, Vector Product Format. Defense Mapping Agency, Fairfax (1989). http://earth-info.nga.mil/publications/specs/printed/600006/600006.pdf
22. Defence Geospatial Information Working Group: DGIWG Feature and Attribute Data Registry. DI JE, London (2011). available from https://www.dgiwg.org/FAD/registers.jsp?register=DFDD
23. Defence Geospatial Information Working Group: Registered Items. DI JE, London (2011). available from https://www.dgiwg.org/dgiwg/htm/registers/registers.htm
24. Defence Geospatial InformationWorking Group: Defence Geospatial Information Framework (DGIF) – Overview (2017) document#200 available from https://www.dgiwg.org/dgiwg-standards
25. NATO Military Committee: MC 296/3, NATO Geospatial Policy, 25 October 2016 (NU)
26. NGA Office of Geomatics: Digital Terrain Elevation Data (DTED) (1990) https://earth-info.nga.mil/index.php?dir=elevation&action=elevation
27. U.S. Department of Defense: Performance Specification Digital Terrain Elevation Data (DTED) MIL-PRF-89020B (2000). http://everyspec.com/MIL-PRF/MIL-PRF-080000-99999/MIL-PRF-89020B_25316/
28. https://igw.mgcp.ws/web/guest/home
29. NGA: Multinational Geospatial Co-production Program (MGCP) (2016). https://s3.amazonaws.com/gpccb/wp-content/uploads/2017/01/10163229/GP_Dellagnello.pdf (Created 12.2016)
30. https://www.arcgis.com/apps/webappviewer/index.html?id=10fc70d81a97452fa232336637d69313
31. NGA: Automated Air Facilities Intelligence Files (National Geospatial Intelligence Agency, Bethesda 1984) available from https://www1.nga.mil/ProductsServices/Aeronautical/AutomatedAirFacilitiesIntelFile/Pages/default.aspx
32. NGA: Digital Vertical Obstruction Files (National Geospatial Intelligence Agency, Bethesda 2008) available from https://www1.nga.mil/ProductsServices/Aeronautical/DigitalVerticalObstructionFile/Pages/default.aspx
33. NGA: Nautical – Hydrographic and Bathymetric Product Descriptions (National Geospatial Intelligence Agency, Bethesda 2009) available from https://www.nga.mil/ProductsServices/NauticalHydrographicBathymetricProduct/Pages/DigitalNauticalChart.aspx

34. NGA: Tactical Ocean Data. https://msix.nga.mil/StaticFiles/Files/TacticalOceanDataPamphlet.pdf
35. DMA: MIL-T-89101: Tactical Pilotage Chart (Defense Mapping Agency, Washington 1995). http://everyspec.com/MIL-SPECS/MIL-SPECS-MIL-T/MIL-T-89101_26055/
36. OGC: OpenGIS City Geography Markup Language (CityGML) Encoding Standard (Open Geospatial Consortium, Wayland 2012) available from http://www.opengeospatial.org/standards/citygml
37. World Meteorological Organization: Introduction to GRIB Edition 1 and GRIB Edition 2 (WMO, Geneva 2003) available from http://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/Introduction_GRIB1-GRIB2.pdf
38. OGC: network Common Data Form (netCDF) standards suite. https://www.opengeospatial.org/standards/netcdf

Gerhard Joos Gerhard Joos studied geodesy in Stuttgart and Calgary and received a doctor's degree from the University of the Bundeswehr Munich. After years of consultancy to utility companies and the Deutsche Bahn, as well as in the defense area. He worked 2006–2009 as Associate Professor at the Technical University of Denmark, Copenhagen. Consecutively he was for 6 years Senior Scientist at the NATO Communication and Information Agency and since 2015 Gerhard Joos teaches as Professor for Geoinformatics at Munich University of Applied Sciences. His research focus is on interoperability, data quality, and GI workflows.

25

26 Geographic Information Systems for Transportation

Knut Jetlund and Bettina Neuhäuser

Contents

K. Jetlund (✉)
Department of Manufacturing and Civil Engineering, Norwegian University of Science and Technology
Gjøvik, Norway
Transport and Society, Norwegian Public Roads Administration
Hamar, Norway
e-mail: knut.jetlund@vegvesen.no

B. Neuhäuser
Verkehrsauskunft Österreich VAO GmbH
Vienna, Austria
e-mail: bettina.neuhaeuser@verkehrsauskunft.at

Abstract

Modern transportation is highly dependent on geospatial information, and the functionality based on the information is so intimately interwoven with our daily lives that its significance is often overlooked until it fails in some way. The use of geospatial information for transportation proceeds from traditional GIS for transport planning and administration of transport networks to the provision of location-based services (LBS) and new applications in intelligent transport systems (ITS), including systems for advanced driver assistance (ADAS) and autonomous driving. This chapter describes the concepts for the use of geospatial information in transportation, starting with the legal framework established by the European Commission to ensure the provision of geospatial information related to transport. The framework is based on European legislations that set out clear obligations for data owners providing information to the public. Furthermore, the chapter describes the theory of navigable digital transport network models, based on the graph theory, and gives examples of databases and services with geospatial information from transport authorities, open data sources and commercial map providers. Location referencing methods used for transportation and standards for information exchange are also described.

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_26

The chapter concludes that the legislations and the datasets and standards available lay a foundation for the further use of geospatial information for transportation and that this use is expected to grow.

Keywords

transportation · ITS · INSPIRE · location referencing · standardization

This chapter addresses the use of GIS for transportation purposes. There is a significant and persistent need for geospatial information and systems in the field of transportation. Administrators, planners, and engineers in transportation organizations around the world depend on geospatial information and systems that support maintenance, analysis, and presentation of the information, and so do road users who depend on route planners, navigation systems, and driver support systems. Geospatial information and location-based services (LBS) are also core components of intelligent transport systems (ITS).

In the very beginning, GIS systems were desktop systems that were mainly applied to help visualize and analyze spatial data for a particular purpose. Nowadays GIS systems have become sophisticated multiuser geodatabases that are merged with other IT databases. Moreover, GIS has become a branch of applied computer science (IT) called geoinformatics, that is using all kinds of communication technologies for various end-user services. This chapter takes into account these broader fields of application but treats the topic more from the perspective of public authorities and end-user services. Special topics such as GIS in logistics and air traffic are not covered. The scope of the chapter is to provide a comprehensive overview of the current use of GIS in transportation and to describe the framework conditions, such as legal requirements and standards for its application.

The structure of this chapter is as follows: First, Sect. 26.1 describes the relationship between GIS and ITS. Next, Sect. 26.2 depicts the overall legal framework related to geospatial information for transportation in Europe. Section 26.3 presents an overview of transport network models, while Sects. 26.4 and 26.5 give details about databases and services for geospatial information for transportation. Section 26.6 provides an overview of location referencing methods used in transportation, and Sect. 26.7 describes relevant standards and specifications for exchanging geospatial information for transportation. Finally, the chapter is wrapped up with the discussion and conclusions in Sect. 26.8.

26.1 GIS and Intelligent Transport Systems

Intelligent transport systems (ITS) comprise telecommunications, automation, and applied computer science and describes the integration of these components into systems or products with traffic-relevant functionality in the context of traffic and transport. Geospatial information technology is a core element for a range of intelligent transport systems. The digital representation and location referencing of the transport network, real-world objects, restrictions, and events is the foundation for legal and safe route planning and navigation. Moreover, the mechanisms for location referencing described in Sect. 26.6 enable georeferencing of information to a location in the real world, in a suitable manner for different uses and situations.

Location-based services (LBS) are mobile services that use position-dependent data to provide the end user with particular information or other types of services. The range of services in the domain of transportation is extensive and includes, for example, services that provide information in real time about accidents, road closures, detours, and traffic flow that are relevant for the road user based on the user's position. Other examples are services that provide information about nearby gas stations, charging points, rest areas, or other transport facilities in the close vicinity of the traveler. In the area of public transport, authorities and transport agencies provide multimodal travel information services to the users about routes, transfer nodes, schedules, and delays based on the user's actual positions during the trip.

Aside from the usual ITS services, there is an emerging area in ITS that is based on digital connectivity and communication between vehicles and transport infrastructure, called cooperative ITS, or for short, C-ITS. C-ITS include location-specific notifications, such as road works, weather conditions, emergency vehicle approaching, and others, that depend on geospatial information and location for both the notification and the receiving vehicle. Furthermore, geospatial information will be a critical component in further development towards fully connected, cooperative, and automated vehicles.

26.2 Overall Legal Framework for Geospatial Information for Transportation in Europe

Geospatial information is required to build up end-user services; therefore, providers of services rely on the availability and accessibility of geospatial information. This circumstance has also been recognized by the European Union (EU), which sets out clear obligations for public and private data holders to make data available and accessible across the European Union. European Directives and supplementing Delegated Regulations define what kind of data must be made available, how the provision must be made, and by whom.

The fact that an increasing range of data needs to be made available has also strengthened the trend towards *Open Data*, which is described in Sect. 26.5.1 more comprehensively. In the following sections, the primary legal frameworks for the provision of relevant geospatial information in transportation are outlined.

26.2.1 INSPIRE

The most important legislation in the area of geospatial information is the Directive 2007/2/EG [1] establishing an infrastructure for spatial information in the European Community, the INSPIRE Directive. It entered into force in 2007. INSPIRE obliges the EU Member States to make available spatial data and spatial data services with the aim of establishing an infrastructure for spatial information across Europe. INSPIRE was initially targeted to areas related to environmental protection and shall support community environmental policies or activities that may have an impact on the environment.

INSPIRE has the aim of harmonizing geospatial information and, consequently, enabling a consistent combination of spatial data by making them interoperable. Interoperability in INSPIRE means the possibility to combine spatial data and services from different sources across the European Union in a consistent way without involving specific efforts of humans or machines. Interoperability may be achieved by either changing (harmonizing) and storing existing datasets or transforming them via services for publication in the INSPIRE infrastructure.

The directive defines metadata and describes in detail different kinds of services to make the data accessible, like discovery, viewing, downloading, transformation, and invoking services. Spatial data owned by or on behalf of government bodies are affected by the directive.

Concerning the content of geospatial information that is included in INSPIRE, there are 34 spatial data themes arranged into three annexes (Annex I, II, and III). These data themes are regarded as essential for the construction of an environmental information system. Besides data like addresses, administrative units, buildings, geology, agriculture, hydrography, land cover and use, and many other ecologically relevant data, also transport networks are a dedicated spatial data theme in INSPIRE. The transport network includes road, rail, air, and water transport networks and related infrastructure, as well as the links between different networks. The transport network theme also includes the Trans-European Transport Network (TEN-T as defined in Decision No 1692/96/EC) [2].

To ensure that the spatial data infrastructures of the EU Member States are compatible and usable in a union, in a transboundary context, the technical implementation of INSPIRE is regulated by a series of implementing rules, which regulate, for example, metadata, data specifications (i.e., data models), network services, and data and service sharing.

All INSPIRE information models are described according to ISO/TC 211 [3–5] standards and implemented in Geography Markup Language (GML) application schemas, and the data is transferred with GML files. GML is the Extensible Markup Language (XML) grammar defined by the Open Geospatial Consortium (OGC) and ISO/TC 211 to express geographical features [6]. The INSPIRE model for Transport Networks [7] is further described in Sect. 26.7.3.

26.2.2 PSI Directive

In Europe, there is a persistent trend towards Open Data (OD), which is data that can be used free of charge with almost no restrictions. This trend can be traced back to Directive 2003/98/EC [8] on the reuse of public sector information. The directive, also known as the PSI Directive, entered into force in 2003 and was revised by the Directive 2013/37/EU [9], which entered into force in 2013. The directive aims to harmonize the different national provisions on the reuse of public sector information in all EU Member States. Open Government Data (OGD), i.e., nonpersonal and noninfrastructure-critical data stocks that are in the public interest, must be made freely accessible for free use, dissemination, and reuse without any restrictions.

Also transport-related information is included in this category and made widely available by public administrations in Europe. Open Government Data (OGD) are usually made available under the Creative Commons Licenses [10], one of several public copyright licenses that enable the free distribution of otherwise copyrighted data. Users are allowed to reproduce and redistribute the material in any format or medium and to edit, modify, or build upon the material for any purpose, including commercial purposes. However, appropriate copyrights and other proprietary notices must be provided, a link to the license must be included, and a statement whether any changes have been made to the original data. Transport-related data provided under the principles of the PSI Directive are frequently referred to as Open Transport Data (OTD). Further details on OGD and OTD are provided in Sect. 26.5.1.

26.2.3 ITS Directive

Alongside the PSI Directive, the Directive 2010/40/EU [11] and its supplementing Delegated Regulations [11], regulate the accessibility, use and reuse of transport and traffic data in the area of intelligent transport systems. The ITS Directive was published in 2010 and is the framework for the deployment of intelligent transport systems in the field of road transport and interfaces with other modes of transport.

The directive authorizes the European Commission to draw up and implement specifications, so-called Delegated Regulations, for the harmonized deployment of ITS services. So far, the following Delegated Regulations have been adopted by the European Commission concerning:

- Harmonized provision for an interoperable EU-wide eCall (EU) No. 305/2013 [12]
- Data and procedures for the provision, where possible, of road safety-related minimum universal traffic information free of charge to users (EU) No. 886/2013 [13]
- The provision of information services for safe and secure parking places for trucks and commercial vehicles (EU) No. 885/2013 [14]
- The provision of EU-wide real-time traffic information services (EU) No. 2015/962 [15]
- The provision of EU-wide multimodal travel information services (EU) No. 2017/1926 [16].

Adopting these Delegated Regulations, the European Commission intends to foster access to data and information in order to stimulate the market for ITS applications. Specific data types that must be made accessible are defined, and obligatory exchange formats and standards are specified. The Delegated Regulations are targeted to the public as well as to private data holders. The data categories set out in Annexes of the Delegated Regulations must be provided via so-called "National Access Points".

The Delegated Regulation (885/2013) [14] aims at a compatible, interoperable, and continuous introduction and application of information services for secure parking for trucks and other commercial vehicles. Private and public car park operators and service providers along the Trans-European Road Network, which primarily comprises the motorway and expressway network, are to make their collected data available in DATEX II [17–20] (Sect. 26.7.8). The described data contents include both static and dynamic data.

According to the Delegated Regulation (2015/962) [15] concerning the provision of real-time traffic information services, road authorities and road operators shall provide the static and dynamic road data they collect and update in a standardized format, if available, or in any other machine-readable format. Static road data are geospatial information, like road network data with a set of attributes, and traffic signs and speed limits, as well as locations of transport infrastructure (for example, tolling stations, charging points, public transport stops). In the field of dynamic road data, the obligatory data exchange format is DATEX II, similar to the Delegated Regulation (885/2013) mentioned above.

The Delegated Regulation (2017/1926) [15] concerning the provision of EU-wide multimodal travel information services states that there are numerous multimodal traveler information services in Europe, but those are primarily limited to their respective EU Member State territory, especially when considering door-to-door routing. Therefore, the Delegated Regulation (2017/1926) defines rules for the accessibility, exchange, and reuse of travel and traffic data and is targeted to transport authorities, transport operators, infrastructure managers, or transport on-demand service providers. This envisaged provision of data shall enable seamless, EU-wide travel information. In Annex I, there is a comprehensive list of data categories that are to be made available over the National Access Points. The defined data categories comprise data that is usually used in multimodal travel information services and routing services. They comprise static data in the area of scheduled timetable data, topography/points of interest, data for route calculation, transport network and infrastructure, and tariff information. Also, dynamic data is recommended to be exchanged like disruptions in all modes of transport, delays or closures of stations or roads, and other data needed to build up a travel information system.

The Delegated Regulation clearly defines the data exchange formats to be used for data provision. In the field of static public transport data, NeTEx (Network Timetable Exchange) and SIRI (Service Interface for Real-Time Information) are the required exchange formats according to the Delegated Regulation. They are described in Sect. 26.7.11 in more detail. Road-related data shall be provided in DATEX II, similar to the other related Delegated Regulations (2017/1926) and (886/2013).

To overcome hurdles and to enable cross-border door-to-door routing via single entry for the end user, the virtual linking of services is considered as a critical solution. The Delegated Regulation defines linking of service as the connection of local, regional, and national travel information systems which are interlinked via application programming interfaces (APIs) to provide routing results based on static and/or dynamic travel and traffic information. The Delegated Regulation recommends that travel information services should be linked by using the OJP (Open Journey Planning) interface that enables decentralized, distributed journey planning [12]. This exchange framework is described more in detail in Sect. 26.7.10.

All the above-mentioned standards and technical specifications are crucial when setting-up harmonized data exchange and interoperable services across the borders and transport operators or transport infrastructure operators.

26.3 Transport Network Models

26.3.1 The Network System

Digital transport networks are the core of transport databases and travel and traffic information services. Transport networks are based on GIS network models that are used for the management or planning of transport networks and also for location-based services (LBS), like routing services in navigation systems. Therefore, a suitable model of the transport network existing in the real world is of utmost importance for these tasks. In this section, a basic overview of transport network modeling shall be provided.

In general, a network can be described as a system of interconnected elements. We can image a transport network as being composed of mainly three parts. The first is the geometric part that represents the transport infrastructure with discrete objects. The second part is the inherent logical model behind the geometric part that describes the relationship and the connection between the single elements of the network. The most common model used for this logical part is the topological data model. The topology enables various spatial analyses in the transport network model. Typical questions that are treated by network analyses concern the quickest way from a start-point to an end-point or are related to distance and vicinity of elements or the accessibility of elements within the transport network [21]. The third part of the transport network is the content related part, the so-called attribute information that holds the relevant describing data related to the individual transport networks elements.

These three parts of the transport network model are described in the following sections in an overview of the essential elements of transport network modeling analysis and the underlying theory underpinning the models. Network modeling is based on fundamental spatial theories, which are tackled in specialized literature that is quoted in the following for further reading and insight.

26.3.2 Graph Theory

The theory of networks is based on graph theory, which enables a discrete mathematical description of networks. According to the theory, graphs are abstract mathematical structures which represent a set of objects together with the connections existing between these objects. The mathematical abstractions of the objects are called vertices (also called nodes), which are used to model pairwise relations between objects. A graph is composed of vertices that are connected by edges (also called arcs or lines) [22].

According to graph theory, vertices are other types of point features. Simple points are zero-dimensional objects that have only the property of location and are stand-alone features. In contrast to a simple point, a vertex is a topological junction representing a shared coordinate pair between intersecting lines and/or polygons. Vertices are defined at each bend along a line or polygon feature that is not the intersection of lines or polygons. So, the vertices enable a logical connection of the elements [21, 23–25].

Graphs may be undirected or directed. In an undirected graph, there is no distinction between the two vertices associated with each edge. This means the edges are composed of unordered pairs of vertices. In contrast to this, a directed graph (or digraph) is directed from one vertex to another. In this case, the graph consists of a set of vertices and a set of arcs, which are composed of ordered pairs of vertices [21, 23–25].

Graph theory and the topological logic in the model are the basis for any spatial analyses in the systems. Specialized literature from authors like *Diestel* [22], *Biggs* et al. [26], *Bondy* and *Murty* [27], *Chartrand* [28], and *Gibbons* [29] describe graph theory in detail.

26.3.3 Geometric Network

One of the essential parts of a transport network is the geometric model. In a GIS system, the user always edits the geometric part of the transport network. In a transport network, these elements are lines that represent the transport infrastructure (roads, railway lines) and points representing crossroads, junctions, or nodes. Transport network models are usually based on vector data models. In general, a vector data model is a representation of the world using points, lines, and polygons. They are suitable to model transport infrastructure because they are conceptualized for storing data that have discrete boundaries, which transport objects indeed have [24].

In particular, a geometric network is a collection of lines (edges) and points (vertices) that are connected along with connectivity rules that are used to represent and model the behavior of common network infrastructure in the real world [21, 23].

The premises in modeling a transport network is that all spatial objects of the traffic network are to be made available in the best possible (highest available) resolution and accuracy. Redundant geometries should be avoided. All spatial objects with topological references to existing objects should be located by references to the existing objects and not by duplicating the geometry. In most cases of transport databases, the model geometry of all transport modes is based on the digitized center lines of the sections. All other geometries can be derived from this fundamental element.

26.3.4 Logical Network (Topology)

When a geometric network is created, the geodatabase also creates a corresponding logical network, which is used to represent and model connectivity relationships between features, a concept that is called topology. In transport network modeling, the connectivity and the spatial relationship between the elements of the network are crucial, in order to model the behavior of common network infrastructure in the real world [21, 23–25].

Topology is a set of rules that model the relationships between nearby points, lines, and polygons and determines how they share coincident geometry. The inclusion of topology into the data model allows, for example, for a single line to represent this shared boundary. Furthermore, topology is also concerned with preserving spatial properties while the network is edited by the user [23–25].

The logical network is managed as a collection of tables that are created and maintained by the geodatabase. These tables record how the features that are involved in a geometric network are connected. Three basic topological principles are the basis for a topological data model:

The connectivity describes the edge-vertex topology for the feature dataset. Every edge has a start-vertex (also called from-vertex) indicating where the edge begins and an end-vertex (also called to-vertex) defining where the edge ends. In between the vertices, there is a line segment, which has its identification number and links to both vertices. This is the basis for many tracing, pathfinding, and routing operations.

The second basic topological principle is area definition or edge-polygon topology. A polygon is defined by a set of edges that are connected so that they are surrounding an area. Each edge is only stored once, which leads to the fact that adjacent polygon boundaries do not overlap [23, 29]. The topological property is stating that line segments connect to surround an area and define a polygon.

Contiguity (the topological property of identifying adjacent polygons by recording the left and right sides of each line segment) is the third topological principle and states that polygons that share a boundary are defined as adjacent. This concept is based on the condition that all edges in a polygon have a direction (refer to edge-vertex topology), which allows determining adjacency information. Polygons that share an edge are defined as contiguous (adjacent) and, therefore, both sides of each edge can be specified. This left and right polygon information is explicitly stored within the attribute information of the topological data model. This is the basis for many neighbors and overlay operations [23, 25, 29].

Topology enables the determination and analysis of spatial relationships of all its included features and is hence the basis for all network analyses. Moreover, the topological model is crucial concerning data quality, as it allows efficient error detection within a vector dataset, as the topology rules need to be kept during editing [30].

26.3.5 Attribute Information

Particular semantic or numerical information describing the character of the network element, the so-called attributes, can be associated with each vector feature [23]. The attributes are used to document the status, conditions, equipment, or information related to the use of the transport elements and allow dedicated analysis of certain aspects of the network. The data attributes of these features are stored in a separate part and are linked via a simple identification number to the geometry.

In order to perform typical network analysis on a transport network, specific network attributes need to be available, like the time to travel a given length of road, which streets are restricted for which vehicles, the speeds along a given road, and which streets are one-way.

Among core attributes like name, data type, and units there are attributes related to the usage type. The usage type specifies how the attribute will be used during the analysis. The most common types of use in GIS systems are either a cost, descriptor, restriction, or hierarchy [29].

The costs are applied to measure and model impedances, such as travel time (transit time on a street) or demand (the volume of garbage picked up on the street). These attributes are divided proportionately along the length of an edge. The cost attribute is particularly important during the calculation of routes. In order to identify the fastest route travel time, attributes for each mode are used, or when identifying the shortest routes, the distance attributes are utilized [29].

Descriptors are attributes that describe characteristics of the network or its elements in an absolute manner. Unlike costs, descriptors are not apportioned, meaning that the value does not depend on the length of the edge element. The number of lanes is an example of a descriptor on a street network. Speed limits or other traffic regulations are other descriptor attributes for a transport network. Although it is not a cost attribute and cannot be used as impedance, it can be used in combination with distance to create a cost attribute (for example, driving time) that can be used as impedance [26].

Besides costs and descriptors, restrictions are common usage types. They can be defined for particular elements, so that passing through a restricted element can be prohibited entirely, avoided, or even preferred during analysis. The most common use of a restriction is to prohibit traversal. Typically, one-way streets can be modeled with a restriction attribute, so they can only be passed through from one end to another but not in the reverse direction. Furthermore, restrictions are used to model mode-specific or vehicle-specific constraints. A restriction attribute on walkways can completely prohibit traversal for other modes. That means that network elements identified as walkways are excluded from any route that is generated for a motorized vehicle [26].

The final typical usage type is a hierarchy that is the order or rank assigned to network elements. Typically, a street network has an attribute on the source features that break the roads down into a discrete number of classes, such as local, secondary, and primary roads. This type of information is also used to set driver preferences for certain road classes in routing [29].

26.4 Transport Network Databases in Operation

Many end-user services, like driver information systems, especially those that provide location information and route guidance instruction, rely on some form of digital geospatial information, i.e., information that represents the location and other properties of earth-bound objects. This demand becomes even more important in a connected mobility system, where vehicles and infrastructure communicate with each other (C-ITS). Besides end-user services, public authorities also rely on digital geospatial information in order to manage and maintain the infrastructure, to manage the traffic on their network, and to manage traffic regulations consistently.

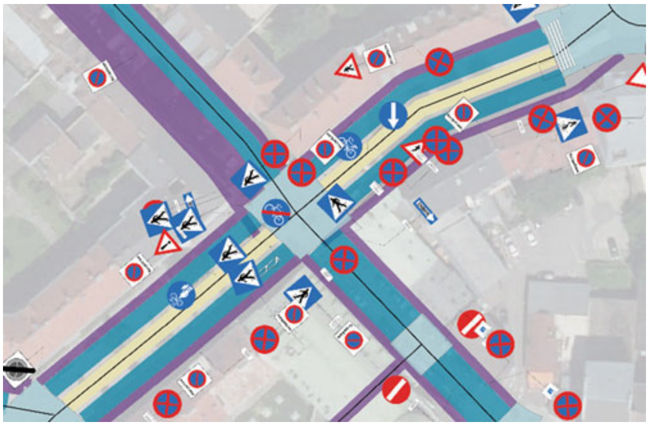


Fig. 26.1 Traffic management and regulation in the Graph Integration Platform © GIP

26.4.1 Road (and Multimodal) Databases

Graph Integration Platform Austria (GIP)

One example of a nation-wide road database is the "Graph Integration Platform Austria (GIP)". The Graph Integration Platform is the multimodal digital transport graph for the whole of Austria. GIP comprises all state, county, and municipal roads, as well as all public transport lines, railway lines, cycling ways, and walking paths. It brings together the various databases and geoinformation systems used to record and manage traffic infrastructure in the public sector throughout Austria. Therefore, it is a common reference graph for all public authorities in Austria.

Aside from the multimodal integration of the transport networks in Austria, GIP is the standard basis to administrate traffic regulations, traffic signs and ground markings. GIP contains permits and prohibitions of the road traffic regulations for all cross-sectional elements (Fig. 26.1). This information is directly connected with the geometric model. Moreover, it includes data on accident black spots and facilities along the traffic network (such as parking garages).

GIP maintains all attributes and usage types that are required for routing. This means that GIP contains a routable node and edge model. All Austrian routes and their connections are represented by edges and nodes (Fig. 26.2). The cross-sectional elements of the road space (e.g., roadway, sidewalk, structurally separate cycle path) are placed on the node and edge model.

For the geocoding of the addresses, the Graph Integration Platform and the official Austrian Address Register were linked. For each address, a reference to the corresponding street section of the Austrian traffic reference graph GIP was automatically created, which has the meaning of an access road to the address. This link enables users of routing systems to obtain more exact information during door-to-door routing. Emergency and rescue services also benefit from the information about the exact access to the address.

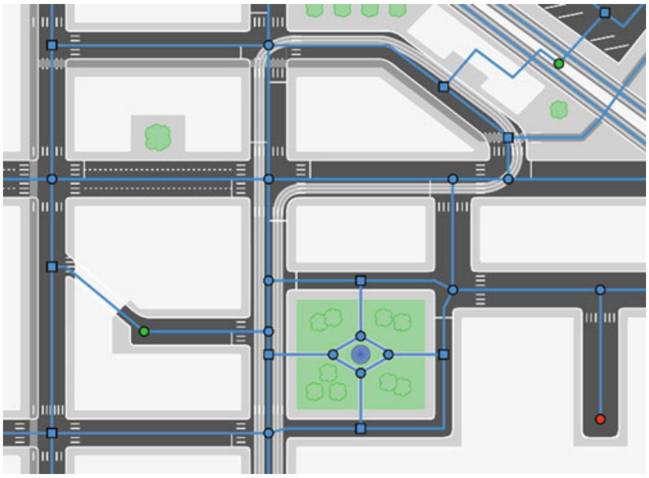


Fig. 26.2 Routable node and edge model © GIP

In the field of public transport, GIP contains all stops and stations of all public transport modes. The current traffic situation is also integrated with the GIP by using real-time data from the road.

Providing all these characteristics and data the GIP is not only used as a basis for traffic and travel information services but also for legally binding administrative and e-government processes, such like administration of roads and paths, and accident data management. Furthermore, GIP is the basis for a cartographic depiction of the transport network, i.e., the Austrian Base Map available as Open Government Data (OGD) [32].

Because it is the official reference graph in Austria, the data quality of GIP has to be secured by agreed codes. In Austria, agreed codes cover the planning, construction, and maintenance of roads, the RVS codes issued by the Austrian Research Association for Roads, Railways, and Transport (FSV). The GIP operation is supported by a specific RVS code (FVS 05.01.14) [33] that comprises a full standard description to be applied for the collection and ongoing

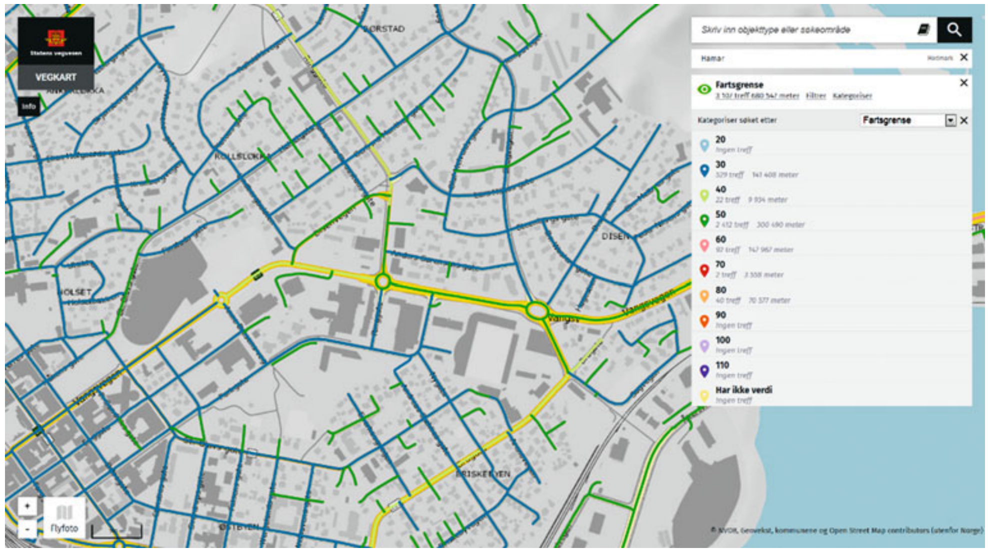


Fig. 26.3 Speed limits presented in Vegkart [31]. Data made available under the Norwegian licence for public data (NLOD), made available by Statens vegvesen

maintenance of the contents of the GIP in order to guarantee the consistency, interoperability, and continuity of the subgraphs necessary for the Austria-wide exchange of traffic references. This ensures that routing, cartographic representations, and basic e-government applications work uniformly throughout Austria.

A series of end-user services are built on the GIP bases. For example, the Austrian Road Operator ASFINAG provides a traffic information service based on the road database and enriches the data by webcam data showing the actual traffic conditions. More than 1000 cameras monitoring the actual traffic status on the main road network [34].

The Norwegian Road Database (NVDB) and the Traffic Information Service

The Norwegian Road Database (NVDB) is the central database for the road network, regulations and restrictions, road equipment, events and other road characteristics in Norway. The database contains geometry and topology for a navigable network for all state, county, municipal, and private roads in Norway, including forestry roads, walkways, and bicycle roads. Altogether, more than 200,000 km of roads are covered. Road properties, such as speed limits, turn restrictions, functional road class, traffic amounts, road width, signs and signals, utilities, accidents, and landslides are located on the road network through linear referencing. Altogether, around 370 different road property types are defined in the feature catalogue for the NVDB.

The road database is the primary information source and storage for the road owners, concerning road development and maintenance, and external users, such as navigation services. For example, navigation datasets and services for emergency vehicles in Norway are based on the digital road network and road properties from NVDB.

The core of the database is the topological network consisting of road links, sequences of road links, and nodes at three topological levels of detail: road level, carriageway level, and lane level. The sequences form the linear referencing system for the database. Each sequence has a fixed length, and both the road links and all road properties are located on the links through interpolated linear referencing.

The road network and road properties in NVDB are publicly available through the NVDB API [35], in JSON and XML format, and through the web client "Vegkart" [31]. The feature catalogue is also available [36]. Figure 26.3 shows how speed limits from NVDB are presented in "Vegkart".

The traffic information service [37] combines static information from NVDB with dynamic information (e.g., traffic flow, road closures, and weather conditions) based on DATEX II services. The service also offers route calculation with information about toll stations and links to road cameras and weather stations.

Figure 26.4 shows the traffic information service with a calculated route, toll stations, traffic flow, and traffic messages. Figure 26.5 shows an example of a road camera and weather information from a mountain road.

26.4.2 Public Transport Databases

The need for digital and integrated geodatabases is also evident in the range of public transport. Such databases are the basis for many operational tasks at a transport operator company. Besides this, they have another crucial function, namely, to offer digital (web or app-based) end-user services

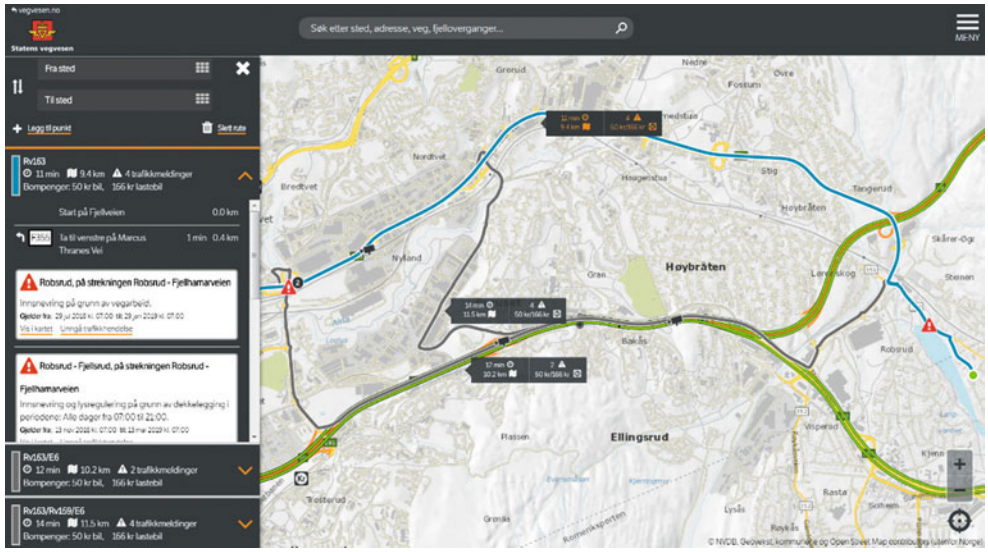


Fig. 26.4 The Norwegian traffic information service [37]. Data made available under the Norwegian licence for public data (NLOD), made available by Statens vegvesen

for travelers, like multimodal travel information and routing services. In this context, public transport databases are the crucial backend of such services.

An example of a public transport database is the Traffic Information Austria (VAO) Platform in Austria. It is composed of an Austrian-wide, public traffic infrastructure geodatabase including the transport networks of all modes in Austria, as well as of an underlying public transport data collection system (Fig. 26.6). In addition to this travel and traffic information for public transport and railways, real-time traffic information from roads is also integrated into this multimodal database.

To build up a common public transport database, it is required to collect and harmonize static public transport data, i.e., stops, travel schedules, and fare data across regional transport associations. The integration of data requires a standard exchange format to be used, like NeTEx (Network Timetable Exchange) [38–40], SIRI (Service Interface for Real-Time Information) [41–45], or national standards like VDV Core Application [41]. In addition to the static data, real-time data on disruption and closures are also collected and subsequently provided within end-user services. These static and dynamic data are linked to the transport network geodatabases, GIP (described in Sect. 26.4.1).

The common multimodal database is made available as a nationwide, multimodal travel information and routing service. In particular, the VAO Platform serves a white-label back-end system and offers access to its service through an XML interface, so that specifically tailored end-user services can be built upon this backend. VAO provides access to data and also to a routing service that cannot only be applied to single modes but also for combinations of modes.

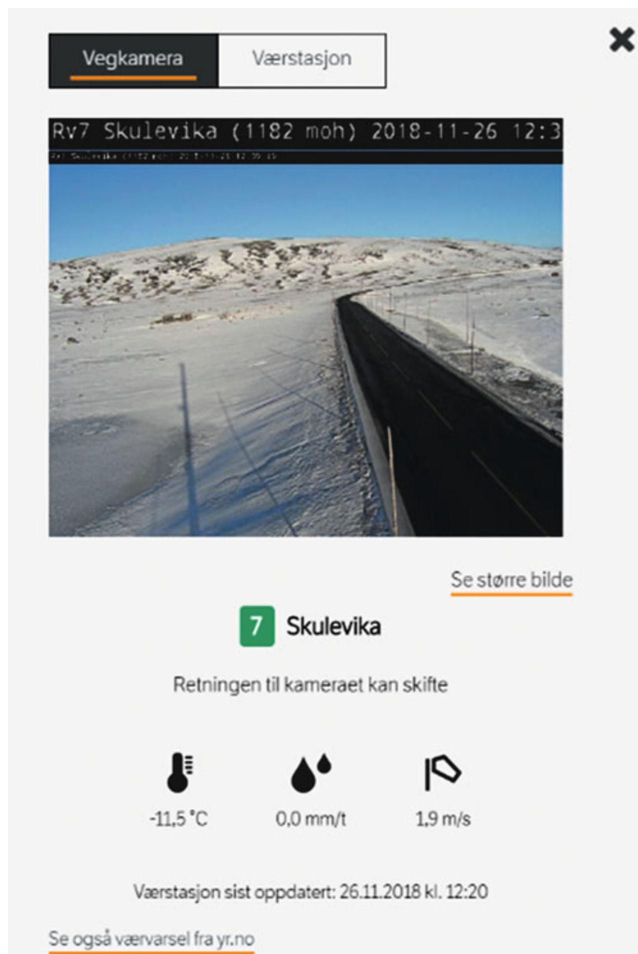


Fig. 26.5 Road camera and weather information from the traffic information service [37]. Data made available under the Norwegian lisence for public data (NLOD), made available by Statens vegvesen

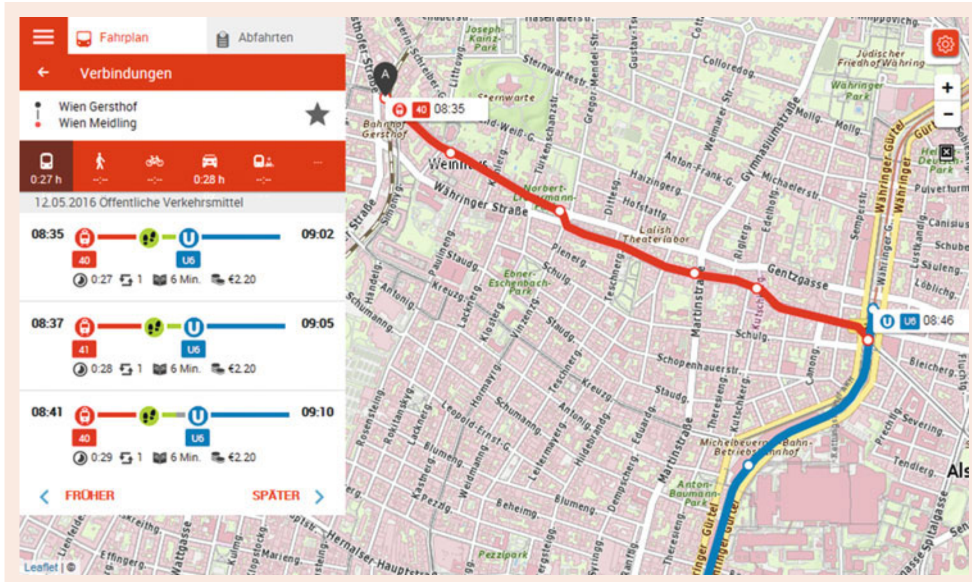


Fig. 26.6 Web-interface for travel information based on VAO Platform © VAO

At the present state, the VAO Platform includes almost 70,000 stops and over 5000 points of interest (POIs). It covers about 8000 km of rail and 2000 km of motorway and expressway network length and around 250,000 km of low-level network.

26.5 Open Data Services

26.5.1 Open Government Data Portals

In recent years, there have been loud calls at a European level and, increasingly, at a national level, for the provision of digital data from the public sector, preferably free of charge and with almost no restrictions on its use as so-called "Open Government Data (OGD)" [46]. These trends are supported on the one hand by EU regulations (Sect. 26.2) and national legislation and, on the other hand, by the steadily growing interest of users of Open Transport Data (OTD) for the sake of transparency in administration, but also due to business opportunities. The European Commission regards OTD as an essential building block for intelligent transport systems to achieve integrated end-user services across modes of transport and Europe through access to data.

Transport Data (TD) include information about traffic infrastructures such as stations, roads, cycle paths, travel times, timetables, digital maps, data on different modes, information on the actual traffic condition (included through real-time sensor data) and data on traffic events (like accidents, construction work, closures). Transport-related statistics, for example on accidents, causalities and safety, and traffic volume and congestion are also included in the transport data (TD) domain.

If these data are made available following the open data principles [46], regardless of whether they are in private or government hands, this is referred to as open transport data (OTD). If they come from the public sector, they are referred to as open government transport data (OGTD).

Data made available as open government transport data (OGTD) are mostly available in the form of CSV (comma separated values), WFS (Web Feature Service), Shapefile, or text-based data formats such as XML and GML. Open data is often also offered in GTFS (General Transit Feed Specification) format and are actively forwarded to Google Maps as a "transit feed".

Regardless of the data format, OGD is made available by the provisions of the PSI Directive [9] and at regular intervals, depending on the organizational strategy. In many European countries, specific metadata services have been established for the provision of OGD data in the sense of visibility and transparency required by the PSI Directive [9]. In Austria, for example, the metadata service, called Open Data Austria [47], provides access to OGD of different domains. It serves as a central catalogue that records the metadata of the decentralized data sources of governmental bodies in Austria and keeps them accessible. Open Data Austria is also the central point of contact to the European Data Portal [48], which collects the open data of all EU Member States as a metadata service. The OGD portal "Felles datakatalog" [49] provides a similar functionality in Norway.

Typical transport-related information that is published as OGD datasets is a road network, public transport network, and stops, but also the real-time data of a transport operator. Based on these open data different location-based services are developed, like location-dependent information on the availability of city bikes or parking spaces and real-time information about the next train or bus at a station.

26.5.2 OpenStreetMap

Besides these legally initiated Open Government Data, there are also private initiatives, like free projects to provide open data. The main free project concerning transport data is OpenStreetMap (OSM), which collects, structures and stores freely usable geodata in a database.

OSM is an open initiative to create and provide free geospatial data. The OpenStreetMap Foundation is a nonprofit organization, a legal body incorporated by the registrar of companies for England and Wales. The goal of the foundation is to support but not control the OpenStreetMap project. OSM collects, structures, and stores freely usable geodata in a database for everyone to use according to the open data principle [46]. These data are under a free license, the Open Data Commons – Open Database License (ODbL) [50]. OSM data is free to use for any purpose, including commercial use.

Data can be downloaded from OSM as an entire dataset or as a selection of smaller areas. The data is usually available as XML formatted OSM files.

26.5.3 Google Transit

Open data are often also offered in the General Transit Feed Specification (GTFS) format [51] and actively forwarded to Google Maps as a "transit feed". GTFS is an industrial standard that defines a digital exchange format for public transport timetables and related geographic information, such as stop locations. GTFS is developed by Google LLC for Google Maps digital map applications, but it is also widely used as an exchange format and to provide OGD.

A GTFS feed includes a series of text files collected in a ZIP file, an archive file format that supports lossless data compression. Each file contains a particular aspect of transit information, like stops, routes, trips, and other schedule data. A public transport operator can produce a GTFS feed to share their public transport information with developers, who write tools that consume GTFS feeds to incorporate public travel information into their applications.

GTFS Realtime is an extension to GTFS and a feed specification that allows public transport operators to provide real-time data. The specification currently supports information on trip updates (like delays, cancellations, changed routes), service alerts (like stop moved, unforeseen events affecting a station, route, or the entire network) and vehicle positions (information about the vehicles including location and congestion level) [51].

26.5.4 TomTom Maps APIs and HERE Location Services

The map providers TomTom [52] and HERE [53] have been delivering datasets for route planning and navigation since the 1980s. Their products now include a range of services for accessing the information by other applications, through TomTom Maps API and HERE Location Services. Examples of such services are search, routing, background maps, and traffic information. Some of these services are free of charge, while payment is required for others.

26.6 Location Referencing Methods Used for Transportation

26.6.1 Location Referencing Concepts

Locations in a GIS are mainly described directly with coordinates referring to a spatial location referencing system, or indirectly by referring to identifiers, such as addresses. In transportation, several other location referencing methods are used, referring to locations in a transport network. Several international and European standards from the GIS and ITS domains describe location referencing methods, and some of these methods are described and compared in the CEN Technical Report on location referencing harmonization for urban ITS [54], while the accompanying Technical Specification [55] describes transformations between the different methods.

According to [54], a location reference is a description of a location within a location referencing system (LRS), according to a location referencing method (LRM). By applying a location reference to a feature such as a transport network link or node, an accident, or a speed limit, the location of the feature is described. The terms "location referencing method" and "location referencing system" are defined in the standard ISO 17572-1 [56]. A "location referencing method" is a "methodology of assigning location references to locations", while a "location referencing system" is a "complete system by which location references are generated, according to a location referencing method".

ISO 17572 specifies two groups of location referencing methods for ITS: precoded location references and dynamic location references. Precoded location referencing is defined as "location reference using a unique identifier that is agreed upon in both sender and receiver system to select a location from a set of precoded locations" and is further specified in part 2 of ISO 17572 [57]. Examples of precoded location referencing methods used for transportation are location by reference to public transport stop points, linear referencing in transport networks, and RDS-TMC messages. Dynamic location referencing is defined as "location reference generated on-the-fly based on geographic properties in a digital map database" and is further specified in part 3 of ISO 17572 [58]. Examples of dynamic location referencing methods are AGORA-C™ [58] and OpenLR™ [67].

26.6.2 Location Referencing by Coordinates

Location referencing by coordinates is the most common method for location referencing in a GIS. The concepts are defined in the standard ISO 19111 [59]. As defined in ISO 19111, a coordinate set that describes a location is a collection of coordinate tuples, where the number of coordinates in the tuple represents the number of axis in the coordinate system. A coordinate tuple that describes a location in two dimensions will have two coordinates (e.g., north and east); while a coordinate tuple for three dimensions will have three coordinates (e.g., north, east, and elevation). A coordinate set for a single point has one coordinate tuple, while a line must have at least two coordinate tuples.

To be able to position the coordinate tuples in the real world, the coordinate set must refer to a coordinate reference system (CRS) that includes a coordinate system (CS) and a datum where the shape of the Earth is defined. A CRS is a specialization of the more general term LRS. The CS (e.g., Cartesian 2-D) includes descriptions of the axis and the unit of measure used for coordinate tuples. The International Association of Oil and Gas Producers (IOGP) maintains a registry of CRSs known as the "EPSG Geodetic Parameter Registry" [60] where every CRS has a unique ID. Some examples of CRSs that are used in GIS for transportation are the World Geodetic System (WGS84) with coordinates in latitude and longitude (EPSG code 4326) for satellite navigation, WGS84/PseudoMercator (EPSG code 3857) for web maps, and ETRS 89 with UTM projections for more accurate positions (e.g., EPSG code 25833 for UTM Zone 33N).

ISO 17572 defines location referencing by coordinates as a dynamic LRM, as it is based on the geographic properties in the dataset and is independent of any other feature such as a transport network.

26.6.3 Location Referencing by Grids

A grid is a division of an area into smaller areas, often referred to as tiles. Each tile can be located within the larger area, typically through coordinates representing the row and column number. The Navigation Data Standard [61] is one example where datasets are divided into grids.

Like location referencing by coordinates, ISO 17572 defines location referencing by grids as a dynamic LRM, as it is based on the geographic properties in the dataset and not on any other feature.

26.6.4 Precoded Location Referencing

Location Referencing by Identifiers

In location referencing by identifiers, a set of identified objects are maintained in a dataset and referred to by their identifiers. The objects must have location references that can be located in the real world, usually coordinates in a CRS. The identified objects must be available and identical for all users of the location references, both those who describe a location (the senders) and those who decode the location (the receivers). A typical example of location referencing by identifiers in GIS for transportation is public transport stop points, such as bus stops, train stations, and airports. These are maintained in datasets owned by transport authorities and may be referred to and identified in a GIS based on their identifiers. For example, the airport codes OSL or MSP are internationally unique and mean Oslo Gardermoen Airport (OSL) and Minneapolis-St Paul International/Wold-Chamberlain Airport (MSP) and are located with WGS 84 latitude longitude in World Airport Codes [62]. At a local level in cities, the name of a bus stop is often used as an identifier and can be located through a dataset from the local transportation company or the local authorities.

The standard ISO 19112 [63] defines GIS concepts for location referencing by geographic identifiers, where the identified objects are called locations and are stored with their real-world position, described with coordinates in a CRS, in a gazetteer.

Linear Referencing

Linear referencing is a more advanced method for precoded location referencing where locations are described on identified elements in a network. Linear referencing is widely used for transport networks; one example is the transport properties in INSPIRE Transport Networks that are located within the network with linear references. The core of linear referencing is a network that forms a linear referencing system, which is a specialization of the more general term LRS. The links in the network must have unique identifiers that can be referred to for objects connected to the network. In addition to this reference, one-dimensional positions along the network links describe locations within a link, either as a single point along the link or as two coordinates describing the start and end of a section of the link. These positions can be given in a measuring unit such as meters along the link geometry; they can be interpolated values between 0 and 1, where 0 represents the start of the link, and 1 represents the end, or they can be measured from some reference point along the link.

By using linear referencing, the core network can be held stable, regardless of changes in characteristics related to the network. This is an advantage for network owners. For example, the network properties in INSPIRE Road Transport Networks [7], such as speed limit and road width, are characteristics that may differ along a road link, while the link itself is unchanged. If these characteristics were attributes to the road links, the links would have to be cut each time one of the attributes changed. With many such characteristics, the result would be a lot of short links and a network that was complicated to maintain. Furthermore, linear referencing is also an effective method for calculating and analyzing combinations of network characteristics, e.g., to identify how many meters of a road network have a given speed limit and road width below a given threshold.

Standard ISO 19148 [64] defines the concepts for linear referencing, with the conceptual model for how linear locations in a network are described. When used in a GIS, the linear references must be transformed to coordinates (point or line geometry) in a CRS for presentation on a map. Advanced GIS applications include tools for transforming linear references to geometry and vice versa.

Location Referencing for RDS-TMC with ALERT-C

The ALERT-C location referencing method is a specific precoded LRM that is defined in ISO 14819-3 [65] to describe locations in traffic messages and, in particular, in RDS-TMC (Radio Data System–Traffic Message Channel). In ALERT-C, locations are referred to predefined locations such as intersections, but information can also be included about direction and extent. For example, a congestion may be located between two intersections, in a specified direction and extending almost up to the last intersection.

Predefined and hierarchically numbered location tables with locations referred to as coordinates in WGS84 latitude and longitude are an essential part of ALERT-C and RDS-TMC. To transform the messages into points and lines for representation in a GIS, it is fundamental to have access to these tables. For example, in Norway, the tables are available in the registry of RDS-TMC services [66].

26.6.5 Dynamic Location Referencing: AGORA-C™ and OpenLR™

With dynamic location referencing methods according to ISO 17572-3 [58], information related to transport networks may be exchanged from a sender to a receiver and matched to the receiver's network dataset, even if the two have different representations of the network. Typically, such differences can be on the topological levels of detail with one or two links representing a bidirectional road.

The location references are based on a simplified geometry representation of the shortest path in the sender's network, accompanied by a set of road characteristics such as form of way and functional road class. This information is decoded on the sender side and encoded and matched to the network on the receiver side.

Two main methods for dynamic location referencing have been developed: AGORA-C™, as described in ISO 17572-3 [58], and OpenLR™ [67], developed by TomTom International. For both methods, the simplified geometry is described with coordinates in latitude and longitude, with a resolution of 10 micro degrees. AGORA-C™ refers to ITRS as the CRS, while OpenLR™ refers to WGS84. For practical terms at this resolution, these can be considered identical.

The encoding process must be performed to enable the transformation of the messages into geometry representation in a GIS. Software for performing the encoding and decoding of OpenLR™ messages is freely available [68], while AGORA-C™ is a licensed method, where licenses can be obtained [69].

26.7 Standards for Provision and Exchange

Transportation of passengers and freight relies on geospatial information from a range of sources, including static and dynamic information about the infrastructure and the surrounding environment (e.g., regulations and restrictions for navigating the infrastructure, weather conditions, traffic amounts, and road closures). Commercial map providers and vehicle manufacturers maintain datasets and services for route planning and navigation, including both static and dynamic information, and share this information with road users in real time. Sensors in vehicles and personal mobile devices collect real-time information about locations (e.g., the nearby physical environment, weather conditions, and road conditions) and share the information with other road users, authorities, and map providers. Infrastructure owners collect and share information such as traffic amounts, average speed, and weather conditions from sensors on and along the transportation networks. Mapping authorities share updated maps of the surrounding environment, and infrastructure owners such as road and rail authorities provide authoritative information about the networks and navigation restrictions, as well as dynamic information (e.g., accidents, closures and detours).

Transportation planners, infrastructure owners, and users of transport systems need to combine, understand, and validate this information into the knowledge needed to plan and perform transportation in a legal, safe, effective, and sustainable manner. For this purpose, standardization and harmonization of information models, services and exchange formats is a crucial element.

26.7.1 Standardization Stakeholders for Geospatial Information Related to Transport

There are many stakeholders involved in standardization activities for geospatial information related to transport, including official international standardization organizations, public authorities, commercial actors, and consortiums.

Official international standards in this domain are mainly developed and maintained by the International Organization for Standardization (ISO), Technical Committees ISO/TC 211 – Geographic Information – Geomatics, and ISO/TC 204 – Intelligent Transportation Systems. ISO/TC 211 has its focus on the core concepts for digital geospatial information in general, including conceptual standards for modeling the real world in a geospatial context, location referencing, exchange formats, and services. ISO/TC 204 works with a broad range of standards for ITS, with Working Group 3 focusing on geospatial information for ITS.

European standards are developed by the European Committee for Standardization (CEN) and European Telecommunications Standards Institute (ETSI), and in particular, the CEN Technical Committee CEN/TC 278 works with geospatial road-related information for use in ITS. In particular, the two working groups WG7 (ITS Spatial data) and WG8 (Road Traffic Data) have worked with the relevant standards for this section. Finally, the work on the INSPIRE data specifications is also essential for geospatial road-related information in Europe.

In addition to official standards, commercial actors and consortiums work with domain-specific standards for transport information. The work of the Open AutoDrive Forum, which is a consortium of ADASIS (Advanced Driver Assistance Systems Interface Specifications), the Navigation Data Standard (NDS) Association, SENSORIS (an industry platform that will develop a global standard for vehicle-to-cloud data), and TISA (Traveler Information Services Association), is of particular importance. For this section, the most relevant organizations in the Open AutoDrive Forum are the NDS Association and TISA. The NDS Association [70] works with specifications of geospatial information needed for automated driving, with a wide range of members from vehicle manufacturers, application developers, map data providers, and service providers. The TISA organization [71] works with specifications for traffic and traveler information and has a list of members from vehicle manufacturers, application developers, service providers, broadcasters, and public authorities. The ADASIS organization works with standards for information exchange for advanced driver assistance systems, while the primary objective of SENSORIS is information exchange with onboard sensors in vehicles.

26.7.2 GDF

The ISO standard ISO 20524-1:2020 Geographic Data Files (GDF), version 5.1 [72] is a standard for exchange of data between databases with geospatial information for road navigation. The scope of GDF is static geospatial information, including the road network with geometry and network topology, restrictions, equipment, and the surrounding environment. The most recent version of GDF covers extended requirements from C-ITS, multimodal transportation and systems for automated driving.

The core of GDF is a generic model with the primary building-blocks feature, relationship and attribute. Features are real-world objects, such as roads, ferries, and road equipment; relationships describe relations between two or more features, such as a maneuver, and attributes describe further characteristics of features or relationships. The standard describes catalogues with specific models of features, relationships, and attributes, based on the basic building blocks.

Implementations of the standard in Media Record and XML formats use the generic model, with the connection to entities from the feature, relationship, and attribute catalogues through identifiers.

Data in GDF files are organized in three levels of representation: Level 0 describes only the geometry of the topological primitive building blocks node, edge, and face. Level 1 describes simple features, for example, road elements or intersections, while Level 2 describes complex features that consist of simple features or other complex features. Furthermore, the topology between features can be described as nonexplicit (no defined topology), nonplanar (defined topology where features may overlap), or planar (defined topology where features must not overlap or intersect each other).

26.7.3 INSPIRE Transport Networks

The INSPIRE Data Specification for Transport Networks [7] describes a core information model with a multimodal network for road, rail, air, and water transport, and specific information models for each transportation mode. The information models are described in UML according to ISO/TC 211 principles and are based on the INSPIRE Generic Network Model [73]. GML schemas for implementation are derived from the models.

The Generic Network Model describes essential network elements, in particular links and nodes, their geometry, and the network topology. Network properties can be connected to the network elements through linear referencing. The network elements and network properties defined in the Generic

Network Model are abstract concepts and are specialized further to implementable concepts in the specification for Transport Networks. The Generic Network Model provides a fundamental shared basis for all kinds of network models touched by INSPIRE (e.g., transport, energy, hydrography). Therefore, the INSPIRE Generic Network Model provides a way to harmonize the geometry of databases, which is essential in the context of static road data exchange.

The INSPIRE Transport Networks part "Common Transport Elements" is designed to be used for all kinds of transport networks. The Common Transport model is a further specialization of the Generic Network Model that is more specific for transportation but independent of transportation modes. The network elements are still abstract and are further specialized in the transportation mode specific models, while some network properties are defined to be used independently of transportation modes, such as traffic flow direction, owner, and maintenance authority, and access restrictions.

The specific models for each transportation mode specialize the network model further for each transportation mode (road, rail, air, and water), into implementable subclasses of network elements and network properties. For example, the road network model contains road links and nodes and road specific network properties such as functional road class, form of way, road name, speed limit, and road width.

26.7.4 TN-ITS

The European specification TN-ITS [74, 75] was initially developed through the ROSATTE [76] and TN-ITS [77] projects, in close collaboration between road authorities and commercial map providers and have been further developed to become a technical specification. The purpose of the projects and the specification was to enable road authorities to deliver changes on static road information (e.g., speed limits and turn restrictions) to map providers, and thereby to enable them to update their datasets. The specification has a generic model with road features that can have road properties. The types of features and properties, and also of values of some property types, is standardized in code lists that are harmonized with the information described by the ITS Directive and the Delegated Regulation 2015/962 and may also be extended with additional values. The road network geometry and topology are not covered by the specification, only the features related to the network.

The location of a road feature based on the TN-ITS specification can be described with several location-referencing methods, where at least one is mandatory. This includes geometry, location referencing by identifiers, linear referencing, and dynamic location referencing (AGORA-C and OpenLR). The linear references are intended to refer to the road network provided by INSPIRE Road Transport Networks.

The TN-ITS specification is based on ISO/TC 211 principles for information modeling and includes a UML model according to ISO 19109 [4], and implementation in GML according to ISO 19136 [6]. It also contains elements imported from other standards, such as the DATEX II model [17–20] for validity conditions and the ISO 14823 (Graphic data dictionary) model for classification of traffic signs [78].

Services providing data through the TN-ITS specification were first implemented in Sweden and Norway as a part of the ROSATTE and TN-ITS projects and were later been implemented in several other European countries, and even more countries are in the process of setting up TN-ITS services through the TN-ITS GO project.

26.7.5 OpenTNF

The OpenTNF specification [79] is an open standard for exchange of transport networks, developed in cooperation between Triona AB, Vianova Systems AS, The Swedish Transport Administration, and the Norwegian Public Roads Administration. It is based on a generic model with links and nodes defining a network, and properties attached to the network through linear referencing. The property types and their structure are defined in an object catalogue as part of the dataset.

With the generic model and the object catalogue, the specification is flexible and may be used to exchange any transport network-related information as long as it is described in the object catalogue. The network model and location referencing model in OpenTNF is conformant to INSPIRE Transport Networks and may be used to exchange INSPIRE compliant data.

OpenTNF may be implemented in any relational database, but SQLite and OGC Geopackage has been preferred and used in actual implementations.

26.7.6 Navigation Data Standard

The Navigation Data Standard describes an information model and compact database storage of information needed for vehicle navigation. The complete standard has restricted access for members only, while the Open Lane Model [61] is an open specification where a limited set of features is available.

The standard is based on 16 building-blocks (Fig. 26.7) used for navigation and map display, visualization, automated driving, and index structure. The Open Lane Model is limited to three core building-blocks for navigation and

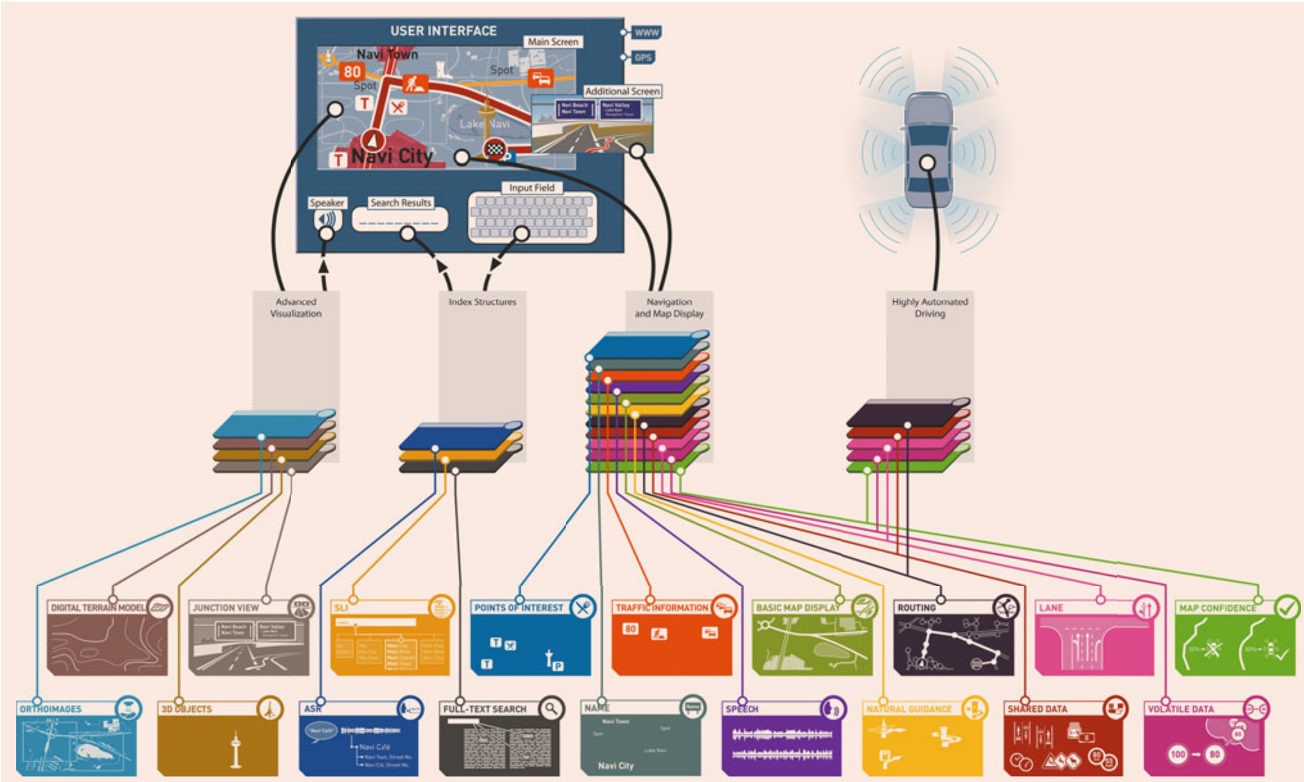


Fig. 26.7 Overview of the NDS building-blocks. (Creator: Emde Grafik [80]. Copyright: NDS e. V. [70] – NDS-Spezifikation CC BY-SA 4.0)

map display: routing (the primary network used for routing), lanes (information needed for navigation at lane-level), and shared data (metadata).

The datasets are stored in SQLite databases according to a generic structure where compact storage and quick access is essential. For this purpose, data are organized in tiles and stored as binary objects in the database.

26.7.7 OpenDRIVE

The OpenDRIVE specification [81] is a data model and an open XML file format for road networks and has mainly been used for simulations. The specification describes a navigable road network based on center lines, with additional information about lanes and road features, including marking, signals, and more. The OpenDRIVE project [81], which developed the specification, has now transferred the responsibility to the Association for Standardization of Automation and Measuring systems (ASAM) for further development and maintenance.

26.7.8 DATEX II

The scope of the DATEX II (CEN EN 16157) series of European standards [82] is dynamic information for traffic and travel, i.e., changes from the static situation for traffic management purposes. Examples can be road closures and detours due to accidents, flooding or landslides, reduced speed due to congestion, and other dynamic information from the road. The standards are explicitly aimed at traffic, infrastructure, and service providers, and contain a standard set of data exchange specifications to enable efficient, seamless, cross-border exchange of traffic data and information. DATEX II has been in operation for over 10 years for national and international traffic data exchange and is currently operated via a network of approximately 95 nodes. Both public and private operators are obliged by the Delegated Regulations [15] to exchange their dynamic and static road data via a national access point using DATEX II or a machine-readable format that is fully compatible and interoperable with DATEX II.

Part 1 of the series describes the main framework and the UML Profile [17]. Parts 3 and 7 describe specific models for situations and publications, including classes with code lists and enumerations for example vehicle classification and validity conditions related to the situations [19, 20]. Each situation is to be described with one of the location referencing methods defined in Part 2 [18], where the possible methods include TPEG, AGORA-C™, OpenLR™, geometry according to ISO 19107, and linear referencing according to ISO 19148.

26.7.9 TPEG2

The scope of the ISO 21219 TPEG2 (Transport Protocol Experts Group) series of standards [83, 84] is transmission of multimodal traffic and travel information, and mainly dynamic information similarly to DATEX II. The standards are developed and maintained by TISA and are provided to ISO/TC 204 for official standardization.

For location referencing of the messages, Part 7 [85] of the series describes the location referencing container, which can be used with several location referencing methods described in other parts, such as coordinates (Part 21) [86] or OpenLR (Part 22) [87].

As there is some degree of overlap between DATEX II and TPEG2, methods for converting from DATEX II to TPEG2 have been developed to enable publication of DATEX II messages using TPEG2.

26.7.10 OJP–Open API

Besides the need for data exchange formats, there is a growing demand for interfaces allowing access to data in a more dynamic way. In particular, in the area of travel information, access to high-quality and up-to-date information is substantial for any reliable end-user service. Therefore, the CEN technical specification specifies an open API for distributed journey planning (CEN/TS 177118:2017) [88]. The technical specification, in short, referred to as OJP (Open Journey Planning API), refers to the concept of linking information services.

In general, APIs (application programming interfaces) are used in web services as an interface to be addressed by other systems. The interface allows the developers to integrate the data provided into their application dynamically. Through an API, data can be exchanged in a structured way so that the data can be further processed by another system and in a reduced form compared to the original data in the source database. The API can grant access to closed data pools and user groups or provide access to a broad group of third parties.

In the field of multimodal travel information, where the OJP applies, this means that by opening an interface and hence linking several travel information systems together, a decentralized, cross-system information system can be established. OJP enables the connection of local, regional, and national travel information systems via technical interfaces with the scope of providing route planning results or other results based on static and/or dynamic travel and traffic information.

The resulting distributed travel planning (routing) provides an effective method to extend the geographical scope of each participating travel planner with minimal effort due to a single standardized API, as defined in [88]. A harmonized interface ensures that only one API needs to be developed for communication between several distributed systems. Such an API can be implemented by any local, regional, or national travel planning system to exchange information with a wide range of other systems.

26.7.11 NeTEx and SIRI

The European standard series NeTEx (Network and Timetable Exchange) [38–40] and SIRI (Service Interface for Real-Time Information) [41–45] are the main exchange frameworks used in public transport. Both NeTEx and SIRI are based on a common basis, namely on the European Public Transport Data Model (Transmodel) (EN 12896:2006) [89], which is the European reference data model. NeTEx and SIRI extract parts of Transmodel, mainly concerning the domain of network topology, timing information (timetables), vehicle scheduling, operations monitoring, and control.

NeTEx is targeted at the exchange of public transport schedules and network data. It enables the exchange of data on passenger information, on the infrastructure and topology of the network (e.g., stops), timetable data, and fare information, as well as on the associated operating data. NeTEx is divided into three parts, each covering a functional subset of the CEN Transmodel for Public Transport Information. The NeTEx XML Schema is based on the concepts and data structures of Transmodel. It is formally drawn up by WG 3 of CEN/TC 278. Since 2014, a series of CEN technical specifications describing the functional scope of the standard has been published (CEN/TS 16614-1:2014 [38], CEN/TS 16614-2:2014 [39], and CEN/TS 16614-3:2015 [40]).

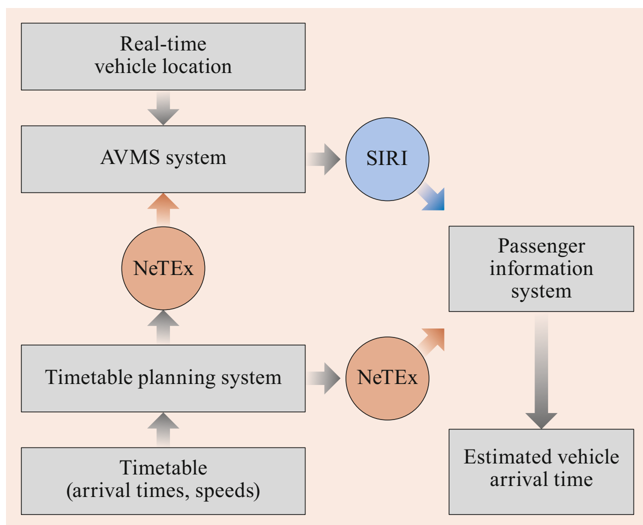


Fig. 26.8 Interplay of NeTEx and SIRI for passenger information (after [90])

While NeTEx is used for the exchange of static planning data, SIRI primarily aims at the exchange of dynamic real-time data. The scope of SIRI is exchanging information about the planned, current, or projected performance of real-time public transport operations. In particular, SIRI facilitates the exchange of structured real-time information about schedules, vehicles, and connections, together with general informational messages related to the operation of the services. SIRI includes CEN technical specifications that are published in functional parts (i.e., CEN/TS 15531-1:2015 [41], CEN/TS 15531-2:2015 [42] CEN/TS 15531-3:2015 [43], CEN/TS 15531-4:2011 [44], and CEN/TS 15531-5:2016 [45]) compliant with CEN Transmodel.

Data in NeTEx and SIRI formats are encoded as XML documents, which must correspond precisely to a standard XML schema. The interplay of NeTEX and SIRI for passenger information is shown in Fig. 26.8.

26.8 Discussion

The insight into the topics provided in this chapter aims to take into account the current state of the use of GIS in transportation. The involvement of GIS in the wide range of applications and standards depicted reflects the current rapid changes in the transportation sector. In the last decades, fundamental changes have been taking place due to new technologies and societal changes. Urbanization and digitalization are among the megatrends that also affect transportation and, subsequently, how we gather, process, and apply data in the area of transportation.

Rapid urbanization requires smart transportation that is effective, well organized, and environmentally friendly. Therefore, traffic management and sophisticated travel and traffic information for citizens are essential. One of the building blocks needed for these purposes is geospatial data. They can support transport infrastructure planning, well-managed operations of the network and vehicles, and even improve traffic safety with real-time information. Therefore, transport network modeling and analysis is a crucial part of knowledge in that area. Transport data need to be collected and harmonized in standard databases, as described in Sects. 26.4 and 26.5, in order to manage a network and the operations on the network efficiently. Provision and reuse of transport data is essential to enable new services.

The building up of integrated geospatial databases supporting traffic management will become even more critical regarding integrated traffic management. While at present state traffic management is carried out for the single modes (road, rail, public transport) separately, integrated traffic management will enable a cross-mode and cross-network management of traffic within the whole mobility system. Also, regarding integrated travel information and integrated ticketing, common databases are in the heart of these developments. Harmonized data and service exchange, and hence standardization of tools, will become increasingly important.

Digitization is at the core of the significant changes taking place in transportation. In particular information and communication technology (ICT) have a massive impact on transportation. New communication technologies, in the area of short-range communication and cellular mobile communication, and new technologies to use big data allow companies to develop advanced smart transportation systems. The enabled connectivity allows the transport industry to develop innovative products and services, like remote sensors delivering real-time information to vehicles, smart traffic lights, self-driving or driverless vehicles, and others. Consequently, the exchange between different components in the mobility system is becoming more and more important. The need to harmonize the communication and the exchange is tackled by standardization and also by European law. Therefore, knowledge on transport-related geospatial information within European standards, directives and delegated acts is crucial.

Summary and Conclusions

In this chapter, we have described concepts for the use of geospatial information in transportation. This use has been extended from desktop GIS systems for administration, planning and route calculation to include location-based services (LBS) and intelligent transport systems (ITS), where geospatial information plays a crucial role.

The European Commission has established a legal framework securing the provision of geospatial information from public authorities and other data owners for use in the transportation sector, through the INSPIRE directive, the PSI directive and the ITS directive with additional delegated regulations. The directives and the delegated regulations include both static and dynamic information and cover multimodal transport modes. Furthermore, they define the data exchange frameworks (European standards or technical specifications) to be used for data provision and exchange.

The core of GIS for transportation is the digital representations of the transport networks, enabled for navigation and route planning. Such networks are based on mathematical descriptions from graph theory and consist of three main parts: the geometry, the topology, and the content. Databases and services with transport networks are maintained and made available to the public; this was exemplified in this chapter with the Graph Integration Platform Austria, Traffic Information Austria, and the Norwegian Road Database. Other datasets for transportation are also available as Open Government Data, through free projects such as OpenStreetMap, and commercial services.

Like any other exchange and use of information, a common understanding of how the information represents the real world is needed. Two particular essential elements to describe and understand geospatial transport information are location referencing methods and standards used for exchanging transport information. In traditional GIS, locations are mainly referenced directly with coordinates in a coordinate reference system, or indirectly with identifiers such as addresses. For transportation, other location referencing methods are also commonly used, in particular, precoded methods such as linear referencing and ALERT-C (RDS-TMC), and dynamic methods such as AGORA-C™ and OpenLR™. Furthermore, several standards and specifications exist and are used. Concerning the legal framework in Europe, the INSPIRE Transport Networks specification, the DATEX II standard, and the PT-Open API are of most importance, while GDF and the Navigation Data Standard are important specifications for geospatial information in the ITS domain.

Transportation is a core part of everyone's daily life, but it is also a significant source of pollution and land consumption. Geospatial information made available through the legislations, datasets and standards available establish a foundation for using geospatial information for transportation and may be used to improve transportation by finding more efficient, safe, and environmentally friendly solutions. It is expected that the use of geospatial information will continue to grow, in particular for location-based services and intelligent transport systems.

References

1. European Parliament and the Council of the European Union: Directive 2007/2/EC of the European Parliament and of the Council of the European Union of 14 March 2007 Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) (2007). EUR-LEX Official Journal of the European Union
2. European Parliament and the Council of the European Union: Decision No 1692/96/EC of the European Parliament and of the Council of 23 July 1996 on Community Guidelines for the Development of the Trans-European Transport Network (1996on). EUR-LEX Official Journal of the European Union
3. ISO 19103:2015: Geographic Information — Conceptual Schema Language. International Organization for Standardization, Geneva (2015)
4. ISO 19109:2015: Geographic Information — Rules for Application Schema. International Organization for Standardization, Geneva (2015)
5. ISO 19131:2007: Geographic Information — Data Product Specifications. International Organization for Standardization, Geneva (2007)
6. ISO 19136-1:2020: Geographic Information — Geography Markup Language (GML) — Part 1: Fundamentals. International Organization for Standardization, Geneva (2020)
7. INSPIRE: INSPIRE Data Specification on Transport Networks – Guidelines (2010)
8. European Parliament and the Council of the European Union: Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the Re-Use of Public Sector Information (2003on). EUR-LEX Official Journal of the European Union
9. European Parliament and the Council of the European Union: Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 Amending Directive 2003/98/EC on the Re-Use of Public Sector Information (2013). EUR-LEX Official Journal of the European Union
10. Creative Commons: The Creative Common License. https://creativecommons.org/licenses/
11. European Parliament and the Council of the European Union: Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010 on the Framework for the Deployment of Intelligent Transport Systems in the Field of Road Transport and for Interfaces with Other Modes of Transport (2010on). EUR-LEX Official Journal of the European Union
12. European Parliament and the Council of the European Union: Commission Delegated Regulation (EU) No 305/2013 of 26 November 2012 Supplementing Directive 2010/40/EU of the European Parliament and of the Council with Regard to the Harmonized Provision for an Interoperable EU-Wide eCall (2012). EUR-LEX Official Journal of the European Union
13. European Parliament and the Council of the European Union: Commission Delegated Regulation (EU) No 886/2013 of 15 May 2013 Supplementing Directive 2010/40/EU of the European Parliament and of the Council (2013of). with Regard to Data and Procedures for the Provision, Where Possible, of Road Safety-Related Minimum Universal Traffic Information free of Charge to User, EUR-LEX Official Journal of the European Union
14. European Parliament and the Council of the European Union: Commission Delegated Regulation (EU) No 885/2013 of 15 May 2013 Supplementing ITS Directive 2010/40/EU of the European Parliament and of the Council (2013of). with Regard to the Provision of Information Services for Safe and Secure Parking Places for Trucks and Commercial Vehicles, EUR-LEX Official Journal of the European Union
15. European Parliament and the Council of the European Union: Commission Delegated Regulation (EU) 2015/962 of 18 December 2014 Supplementing Directive 2010/40/EU of the European Parliament and of the Council with Regard to the Provision of EU-Wide Real-Time Traffic Information Services (2014). EUR-LEX Official Journal of the European Union
16. European Parliament and the Council of the European Union: Commission Delegated Regulation (EU) 2017/1926 of 31 May 2017 Supplementing Directive 2010/40/EU of the European Parliament and of the Council with Regard to the Provision of EU-Wide Multimodal Travel Information Services (2017). EUR-LEX Official Journal of the European Union
17. EN 16157-1:2018: Intelligent Transport Systems – DATEX II Data Exchange Specifications for Traffic Management and Information – Part 1: Context and Framework (2018)
18. FprEN 16157-2: Intelligent Transport Systems – DATEX II Data Exchange Specifications for Traffic Management and Information – Part 2: Location Referencing (2018)
19. EN 16157-3:2018: Intelligent Transport Systems – DATEX II Data Exchange Specifications for Traffic Management and Information – Part 3: Situation Publication (2018)
20. EN 16157-7:2018: Intelligent Transport Systems – DATEX II Data Exchange Specifications for Traffic Management and Information – Part 7: Common Data Elements (2018)
21. Davis, B.E.: GIS: A Visual Approach, 2nd edn. Onword Press, Thomson Learning, Albany (2001)
22. Diestel, R.: Graphentheorie, 3rd edn. Springer, Heidelberg (2005)
23. Campbell, J.E., Shin, M.: Essentials of Geographic Information Systems. Flat World Knowledge, Inc., Boston (2017)

24. Aronoff, S.: Geographic Information Systems: A Management Perspective. WDL Publications, Ottawa (1989)
25. Shekhar, S., Xiong, H.: Encyclopedia of GIS, 2nd edn. Springer, Heidelberg (2008)
26. Biggs, N., Lloyd, E.K., Wilson, R.J.: Graph Theory, 1736–1936. Oxford Univ. Press, Oxford (1986)
27. Bondy, J.A., Murty, U.S.R.: Graph Theory. Graduate Texts in Mathematics. Springer, Heidelberg (2008)
28. Chartrand, G.: Introductory Graph Theory. Dover Publications, Mineola (1985)
29. Gibbons, A.: Algorithmic Graph Theory. Cambridge Univ. Press, Cambridge (1985)
30. Zeiler, M.: Modeling Our World: The ESRI Guide to Geodatabase Design. ESRI Press, New York (1999)
31. Statens vegvesen: Vegkart. http://www.vegkart.no/
32. Stadt Wien und Österreichische Länder: Verwaltungsgrundkarte von Österreich. https://www.basemap.at/
33. Austrian Research Association for Road, Railways and Transport: RVS 05.01.14 Intermodaler Verkehrsgraph Österreich – Standardbeschreibung GIP (Graphenintegrationsplattform) – Grundtext, Guidelines and Regulations for the Planning, Construction and Maintenance of Roadways. FSV, Vienna (2014)
34. ASFINAG: ASFINAG Webcams. https://www.asfinag.at/verkehr/webcams/
35. Statens vegvesen: NVDB API. http://api.vegdata.no/
36. Statens vegvesen: NVDB Datakatalogen. https://datakatalogen.vegdata.no/
37. Statens vegvesen: Trafikkinformasjon. https://www.vegvesen.no/trafikk
38. CEN/TS 16614-1:2014: Public Transport – Network and Timetable Exchange (NeTEx) – Part 1: Public Transport Network Topology Exchange Format (2014)
39. CEN/TS 16614-2:2014: Public Transport – Network and Timetable Exchange (NeTEx) – Part 2: Public Transport Scheduled Timetables Exchange Format (2014)
40. CEN/TS 16614-3:2016: Public Transport – Network and Timetable Exchange (NeTEx) – Part 3: Public Transport Fares Exchange Format (2016)
41. EN 15531-1:2015: Standard Interface for Real-Time Information – Part 1: Context and Framework (2015)
42. EN 15531-2:2015: Standard Interface for Real-Time Information – Part 2: Communications Infrastructure (2015)
43. EN 15531-3:2015: Standard Interface for Real-Time Information – Part 3: Functional Service Interfaces (2015)
44. CEN/TS 15531-4:2011: Standard Interface for Real-Time Information – Part 4: Functional Service Interfaces – Facility Monitoring (2011)
45. CEN/TS 15531-5:2016: Standard Interface for Real-Time Information – Part 5: Functional Service Interfaces – Situation Exchange (2016)
46. Open Definition project: Open Definition 2.1. https://opendefinition.org/od/2.1/en/
47. Open Data Austria. https://www.data.gv.at/
48. Publications Office of the European Union: The official portal for European data. https://www.europeandataportal.eu/
49. Felles datakatalog. https://data.norge.no/
50. Open Data Commons: Open Data Commons Open Database License (ODbL). https://opendatacommons.org/licenses/odbl/index.html
51. General Transit Feed Specification. http://gtfs.org
52. TomTom International, B.V. https://www.tomtom.com/
53. HERE. https://www.here.com/
54. CEN/TR 17297-1:2019: Intelligent Transport Systems — Location Referencing Harmonisation for Urban-ITS — State of the Art and Guidelines (2019)
55. CEN/TS 17297-2:2019: Intelligent Transport Systems — Location Referencing Harmonisation for Urban-ITS — Part 2: Translation Methods (2019)
56. ISO 17572-1:2015: Intelligent Transport System (ITS) — Location Referencing for Geographic Databases — Part 1: General Requirements and Conceptual Model. International Organization for Standardization, Geneva (2015)
57. ISO 17572-2:2015: Intelligent Transport System (ITS) — Location Referencing for Geographic Databases — Part 2: Precoded Location References (Precoded Profile). International Organization for Standardization, Geneva (2015)
58. ISO 17572-3:2015: Intelligent Transport System (ITS) — Location Referencing for Geographic Databases — Part 3: Dynamic Location References (Dynamic Profile). International Organization for Standardization, Geneva (2015)
59. ISO 19111:2019: Geographic Information — Referencing by Coordinates. International Organization for Standardization, Geneva (2019)
60. International Association of Oil and Gas Producers: EPSG Geodetic Parameter Registry. https://www.epsg-registry.org
61. The Navigation Data Standard: Open Lane Model Documentation, http://www.openlanemodel.org/ (2016)
62. World Airport Codes. https://www.world-airport-codes.com
63. ISO 19112:2019: Geographic information – Spatial Referencing by Geographic Identifiers. International Organization for Standardization, Geneva (2019)
64. ISO 19148:2021: Geographic Information — Linear Referencing. International Organization for Standardization, Geneva (2021)
65. ISO 14819-3:2013: Intelligent Transport Systems – Traffic and Travel Information Messages via Traffic Message Coding – Part 3: Location Referencing for Radio Data System – Traffic Message Channel (RDS-TMC) Using ALERT-C. International Organization for Standardization, Geneva (2013)
66. Statens vegvesen: RDS-TMC services in Norway. https://www.vegvesen.no/en/professional/trafikk/rds-tmc
67. TomTom International, B.V.: OpenLR Whitepaper Version: 1.5 revision 2 (2012)
68. TomTom International, B.V.: OpenLR. http://openlr.org
69. Via Group: Via Group Licensing. https://www.via-corp.com/licensing/agora-c/
70. Navigation Data Standard (NDS) e.V.: NDS Association. https://nds-association.org/
71. Travel Information Services Association (TISA). https://tisa.org/
72. ISO 20524-1:2020: Intelligent Transport Systems – Geographic Data Files (GDF) – GDF 5.1. International Organization for Standardization, Geneva (2020)
73. INSPIRE: INSPIRE Data Specifications – Base Models – Generic Network Model (2013)
74. TN-ITS. https://tn-its.eu/
75. CEN/TS 17268:2018: Intelligent Transport Systems — ITS Spatial Data — Data Exchange on Changes in Road Attributes (2018)
76. Wikström, L., Landwehr, M., Bock, H., Nasr, A., Wevers, K., Svensk, P.-O., T'Siobbel, S., Boterbergh, B., Blaive, L., Hamish, K., Isakson, P., Bergman, O., Haspe, U., Schlützle, R.: ROSATTE: ROad Safety ATTributes Exchange infrastructure in Europe: Deliverable D3.1: Specification of data exchange methods (2009). Tech. Rep.
77. Borzacchiello, M.T., Boguslawski, R., Pignatelli, F.: Improving Accuracy in Road Safety Data Exchange for Navigation Systems: European Union Location Framework Transportation Pilot (2016). Tech. Rep. 1831–9424, http://publications.jrc.ec.europa.eu/repository/handle/JRC104569
78. ISO 14823:2017: Intelligent Transport Systems – Graphic Data Dictionary. International Organization for Standardization, Geneva (2017)

79. Triona AB: OpenTNF White paper Version 1.0, https://github.com/OpenTNF/opentnf (2017)
80. Emde Grafik. https://www.emde-grafik.de/
81. OpenDRIVE: OpenDRIVE Format Specification, Rev. 1.4. VIRES Simulationstechnologie GmbH, Bad Aibling (2015), http://www.opendrive.org/
82. DATEX II (CEN EN 16157) https://datex2.eu/
83. TISA Executive Office: TPEG – What is it all about? (2014)
84. Transport Protocol Experts Group (TPEG) https://tisa.org/technologies/tpeg/
85. ISO/TS 21219-7:2017: Intelligent Transport Systems — Traffic and Travel Information (TTI) via Transport Protocol Experts Group, Generation 2 (TPEG2) — Part 7: Location Referencing Container (TPEG2-LRC). International Organization for Standardization, Geneva (2017)
86. ISO/TS 21219-21:2018: Intelligent Transport Systems — Traffic and Travel Information via Transport Protocol Experts Group, Generation 2 (TPEG2) — Part 21: Geographic Location Referencing (TPEG-GLR). International Organization for Standardization, Geneva (2018)
87. ISO/TS 21219-22:2017: Intelligent Transport Systems — Traffic and Travel Information (TTI) via Transport Protocol Experts Group (TPEG2) — Part 22: OpenLR Location Referencing (TPEG2-OLR). International Organization for Standardization, Geneva (2017)
88. CEN/TS 17118:2017: Public Transport – Open API for Distributed Journey Planning (2017)
89. EN 12896:2006: Public Transport Reference Data Model ("Transmodel v5.1") (2006)
90. Association of German Transport Companies (VDV): Standardisierter Austausch von Liniennetz- und Fahrplandaten mit der europäischen Norm NeTEx. NeTEx Application Handbook, VDV No. 462 (2020)

Knut Jetlund Knut Jetlund holds a Master's in Technology in Geomatics from the Norwegian University of Life Sciences and a PhD in Engineering from the Norwegian University of Science and Technology. He works for the Norwegian Public Roads Administration as a specialist in standardization of geospatial information across the application domains of Geographic Information Systems, Intelligent Transport Systems and Building Information Modelling.

Bettina Neuhäuser Bettina Neuhäuser worked at AustriaTech, an agency of the Federal Ministry of Transport, Innovation and Technology, focusing on international deployment projects in the field of Intelligent Transport Systems. She obtained a research fellowship from the Austrian Academy of Science and received her PhD in the area of Geoinformation from the Julius Maximilians University of Würzburg in 2010. She now works as service manager at VAO GmbH, providing multimodal travel and traffic information services for Austria.

26

Geology

27

Kristine Asch, Jens Klump, Stephen J. Mathers, and Holger Kessler

Contents

K. Asch (✉)
Bundesanstalt für Geowissenschaften und Rohstoffe
Hannover, Germany
e-mail: kristine.asch@bgr.de

J. Klump
Mineral Resources, CSIRO
6151 Perth, Australia
e-mail: jens.klump@csiro.au

S. J. Mathers
e-mail: steve.mathers35a@gmail.com

H. Kessler
Geological Modelling Systems, British Geological Survey
Nottingham, UK
e-mail: hke@bgs.ac.uk

Abstract

This chapter explains the historic impact and future direction of geoinformatics in the geological sciences. The history and the purpose of geological surveys are discussed and an introduction to the basic techniques of traditional field mapping is presented. The authors then discuss the recent emergence of digital field capture tools and 3-D geological modeling software and methodologies, which are beginning to replace 2-D techniques. The main impact of these advances is that geologists are now able to capture their knowledge in digital 3-D form, freeing them from the constraints of 2-D media such as paper and (later) GIS. The delivery of geoscience information through 3-D viewers and over the web is going to revolutionize the way in which geologists are able to communicate their science. Perspectives on why geology is only a small (but important) part of the wider environmental science community and how the need for whole Earth system science is forcing geological survey organizations to cooperate more closely with other disciplines are provided. Finally, this chapter details several global initiatives that have gone a long way to achieving global agreements on standards, institutional arrangements, and policies to enable geoscience information to be accessible across discipline and political boundaries. Facilitating the conceptual and computational integration and interoperability of static models (e. g., those focusing on geology and infrastructure) with dynamic models (e. g., groundwater and flood forecast models) to provide decision-makers with science-based decision-making tools will be the main challenge in the field of geoinformatics for years to come.

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_27

Keywords

geology · spatial data · 3-D models · environmental and subsurface management · spatial data instructions · INSPIRE

Geology is the interface between Earth and its resources. For centuries, geological maps of any kind have been the common visualization tools for geology. Geological maps not only map the geology at the Earth's surface but they also interpret the 3-D structure of the Earth's crust, with additional insights coming from outcrops, road and canal cuttings, boreholes, and mine shafts. Examples of early geological maps include a petrographic map of the Electorate of Saxony and the Incorporated Countries [1], a map of the entire Transbaikal region produced by *Lebedev and Ivanov* (1788–1798), and a geological map of England and Wales produced by *William "Strata" Smith,* which established the principles of chronostratigraphy and biostratigraphy [2, 3].

Over the last few centuries, geological maps have undergone considerable development from mostly black-and-white rudimentary displays of natural resources bearing strata through graphically intricate and beautifully colored paper maps (Figs. 27.1, 27.2) to 3-D and 4-D models available on the Internet. Geographic information systems (GIS) and web mapping allow the storage, retrieval, and presentation of far more information and knowledge about a specific area than a 2-D paper map, so the paper map has been superseded by electronic media, even though survey mapping areas are still referred to as *map sheets*.

Geologists can, therefore, be said to have developed a conceptual model—a formalized mental image giving a simplified view of relevant aspects in the real world. The full conceptual model only exists in the mind of the author and must be further simplified for representation on the map. The limitations of our tools force us to reduce the complexity of nature to a few mappable units. The completed map is a permanent, shareable public record of the author's ideas. However, it imposes physical constraints on the representation of the conceptual model [5].

The geological map has been the means for geologists to record, store, and disseminate their knowledge and the results of their investigations of the rocks and unconsolidated deposits of the Earth's surface. Geological maps usually simply show what is where on the Earth's surface [6]. They do, however, often also display the distribution of geological units under the surface in so-called uncovered maps, where the overlying strata are virtually removed and the underlying geology is shown according to reconstructions from combined outcrop and field/borehole information, geophysical investigations, and interpretation. Furthermore, geological maps group geological units into a stratigraphy that places them in genetic and temporal context.

An understanding of geology is essential for protecting human life, health, and assets and for sustaining our environment and resources, and geological maps have always

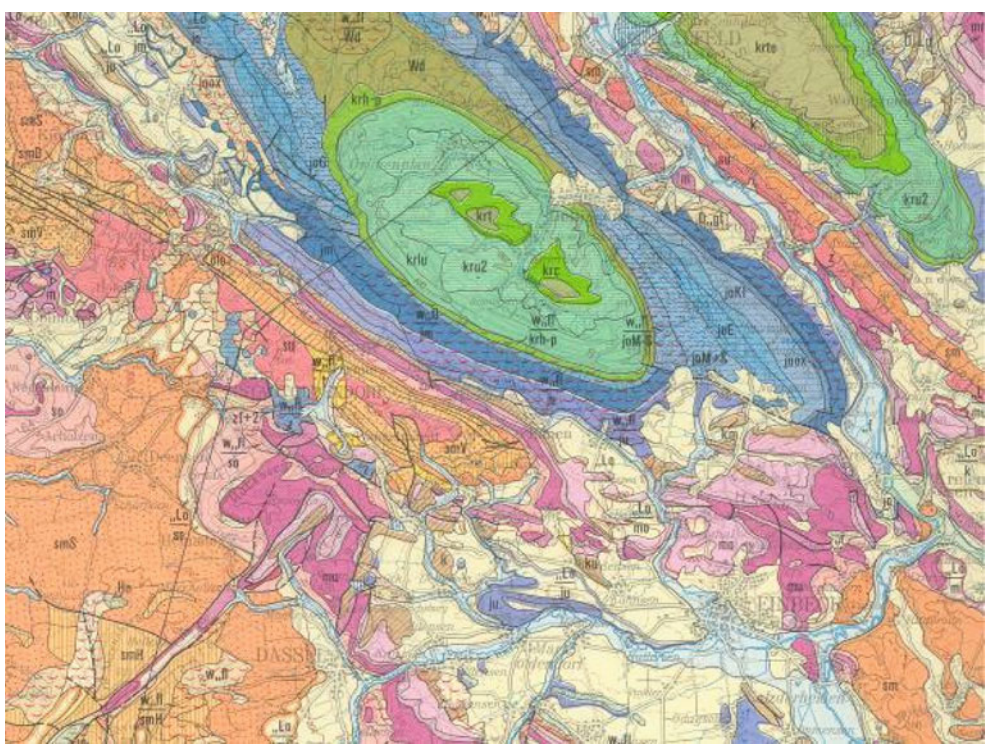


Fig. 27.1 Detail from a classical geological map (from map sheet Kassel of the 1 : 200,000 Geological Map of Germany (GÜK)) from *Zitzmann* and *Lepper* [4]. The *colors* depict the age of the rock, the rock type is indicated by the type of *hatching* used, and the *symbols* explain the rock age and overlaying structures in detail [4]. Topographic base: TÜK200, CC4718 Kassel, Institut für Angewandte Geodäsie 1977

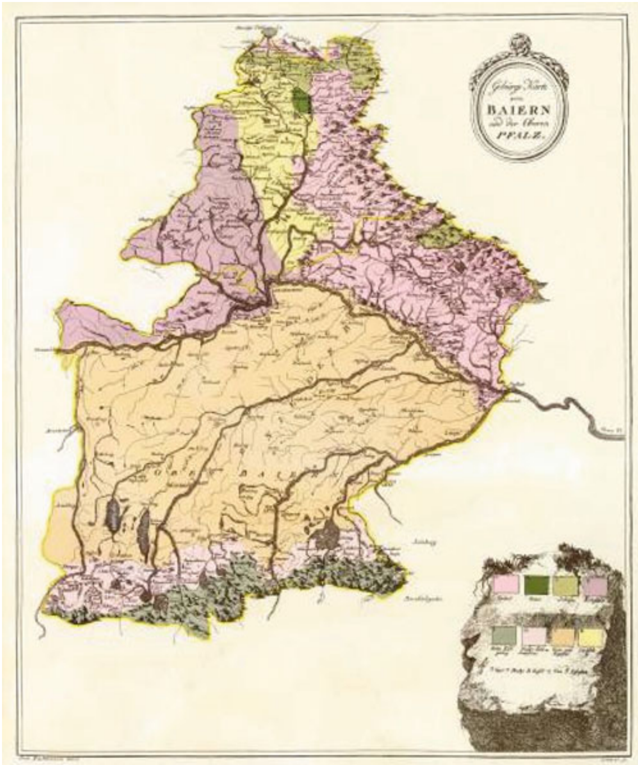


Fig. 27.2 Historical geological map of Bavaria and the Upper Palatinate

provided basic knowledge about the distribution of natural resources: they are the basis for understanding the Earth and its processes. Geological maps may, albeit indirectly, warn about the dangers of natural hazards or supply information about suitable sites for landfills, house building, or tourism. They provide the basis for the commercial exploration of mineral and energy resources and for environmental planning and protection, and they support public-policy decision making [7, 8]. As in many areas of life today, information technology (IT) has a significant impact on the way geological map data and knowledge are captured, processed, and disseminated. It is no exaggeration to say that the effective application of IT is the key to the future exploitation of geological knowledge for the benefit of society.

To the expert eye, geological map data not only depict the surface distribution of rocks but also predict their disposition at depth and, therefore, provide a glimpse into the third dimension. However, even with the sophisticated information technology of today, it is still not feasible to automatically generate a definitive, all-encompassing three-dimensional model of what is beneath the Earth's surface. As the Earth is not transparent, it needs the interpretation skills and knowledge of a geologist to elucidate the world's subsurface from the evidence provided by field mapping, boreholes, geophysics, etc. Thus, a never-completed 3-D or 4-D model of the small piece of crust being surveyed is developed in the geologist's mind, looking behind the map. The geological map—with its typical marginalia, the legend, the horizontal and vertical sections, and block diagrams—is the way that geologists transfer their unique understanding to the reader.

Geological maps, produced by the partnership of skilled cartographers and geologists, exemplify the art of presenting complex science in an aesthetically pleasing way. These skills were taken forward and refined during the eighteenth and nineteenth centuries in several countries and often in parallel. These maps frequently attempted to hold and display every detail and piece of knowledge the scientists had recorded.

With the advent of IT, however, some former constraints no longer exist. Modern computing systems (for example, databases, GIS, and Internet tools) allow the storage, retrieval, and presentation of far more information and knowledge about an area than a 2-D piece of paper. The key point is the separation of the storage and recording of information from its means of dissemination; we are no longer forced to serve all purposes with the same general-purpose document. Using IT, we can select the area, change the scale and the topographic background, select the theme, and amend the colors and line styles. We can distribute the knowledge in many ways, delivering it on paper, on digital media, or across the web, and choose a variety of resolutions, qualities, and levels of complexity. Increasingly, geologists have started to use modeling software to create 3-D and 4-D models, allowing users, through a variety of visualization methods, an insight into the original scientist's interpretation of the Earth below our feet.

27.1 Field Work

Mapping methods used in the field have changed enormously in the last decade: classical geological field data collection, analysis, and map compilation have undergone and are still undergoing fast modernization of methods, mainly driven by the development of global navigation satellite systems (GNSS) such as the Global Positioning System (GPS) and GIS, as well as advances in high-resolution digital aerial imagery and GIS-equipped mobile computers. Whereas a geologist would have previously used her or his geologic compass, field book, paper field maps, and hammer, modern field mapping looks considerably different: handheld personal digital assistants (PDAs) and portable PCs that are equipped with GNSS and GIS, especially stable, and constructed for field work record a wide spectrum of geological data and facilitate iterative in situ digital geological map construction. These data can then be presented in a variety of formats, e. g., as vector or raster data maps in different projections with different topographic bases, and can be published via online mapping services.

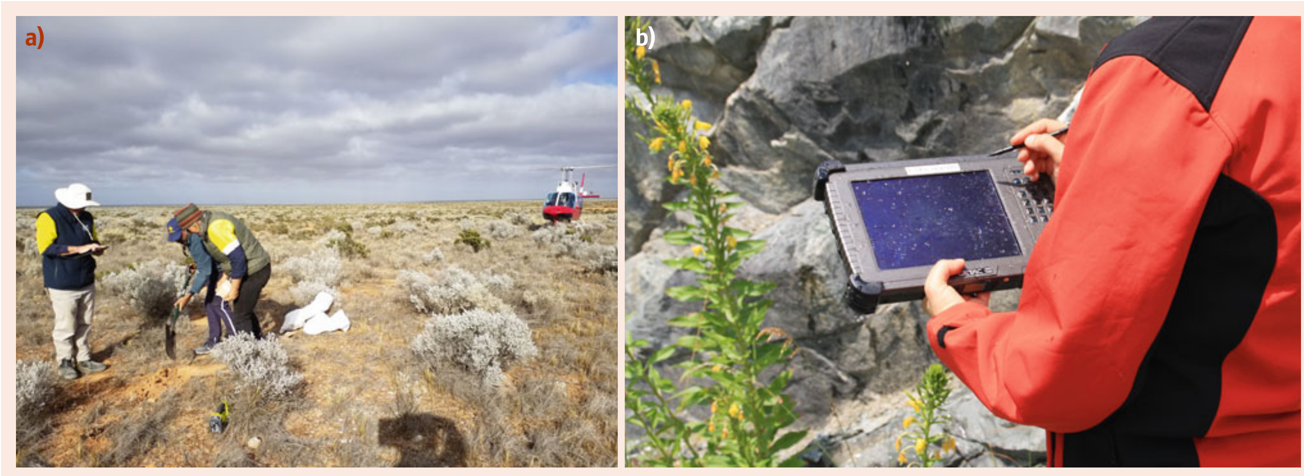


Fig. 27.3 **a** Geological mapping with a mobile device (with GNSS capability) in the Nullarbor Desert of South Australia (from [9] Noble et al. 2018). **b** Modern geological mapping with a weatherproof rugged laptop. Source: Bayerisches Geologisches Landesamt für Umwelt; photo by G. Loth

Geological mapping is based on direct observations from outcrops and geomorphology, rock sample collection (Fig. 27.3), and borehole data analysis. The result is always an interpretation of the subsurface that considers the age, the lithology, and the genesis of the rock units and their structures. Mineralogical and geochemical analyses of the collected rocks and borehole samples help to refine the results and their interpretations, and are excellently complemented by geophysical observations using the radioactivity, resistivity, and electromagnetic or acoustic properties of the rocks. Geophysical methods comprise electromagnetic and georadar measurements, gravity and acoustic techniques, magnetic anomaly mapping, and seismic methods.

27.2 Geographic Information in Geology

27.2.1 Influence of Geographic Information on Geology

Information technology has been used within the geosciences for a long time. In the late 1960s and early 1970s, computing was virtually restricted to those geoscientific disciplines that required significant computational power, i. e., geophysics and geochemistry. A few pioneers in the 1970s [10] tried to apply IT to the production and manipulation of *graphic* geological map data but had limited success. The technology and processing power proved to be insufficient to achieve their goals. By the mid-1980s, graphic computing functionality became widely available in commercial software packages.

Computer-aided design (CAD) and relational database systems matured, and several geological survey organizations began to exploit these technologies for geological map production. CAD systems were originally designed for engineering applications. They are vector-based and use points, lines (i. e., connected strings of points), and symbols to build a drawing or a map. Areas are shown by lines enclosing a polygon. Some CAD systems were specifically developed for cartographic purposes, e. g., Bentley's MicroStation and Autodesk AutoCAD Map. These systems became progressively better at representing maps digitally and configuring offset printing and plotting. By the early 1990s, digital cartographic systems had become commonplace and a new technology was beginning to mature: geographic information systems (GIS). The use of GIS as a tool to store, process, analyze, and distribute spatial geological information is now commonplace. Applications of GIS stretch from regional and environmental mapping, resource assessment, remote sensing, soil sciences, and engineering geology to hazard mapping and hydrogeological and marine mapping. With the advent of XML (Extensible Markup Language) and its derivatives (see GeoSciML; EarthResourceML), data exchange and distributed data manipulation (e. g., in OneGeology and OneGeology-Europe) became much easier.

27.2.2 GI Standards in Geology

Interoperability in geology is motivated by the need to respond to an increased societal demand for geoscience information and by the concurrent opportunity to leverage technological advancements in processing heterogeneous and distributed data. Indeed, the latter is often considered an essential part of any approach to opening up formerly highly integrated geoscience data to extended societal applications. This is largely due to the fact that geoscience data are highly specialized and must be transferred into modes

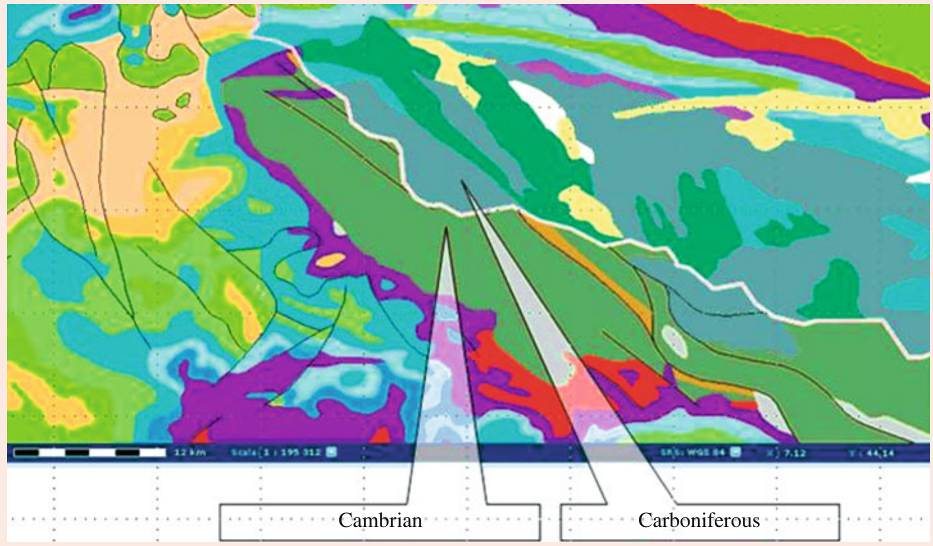


Fig. 27.4 An example of mismatched geological units across political boundaries (Italy/France near Cuneo/Colmar). The geological age is described differently by different national geological survey organizations (after [11])

that are usable by other disciplines and sectors. Such a transformation inherently relies on the integration of data from multiple geology data providers. Until not too long ago valuable information was only available at local, regional, and national levels. Thus, this information was, and in many cases still is, difficult to exploit in the EU or international context due to the high costs associated with its integration. This has led to the information being overlooked in decision making, and has increased the need to develop geoscientific interoperability solutions. The prevalent approaches to institutionalized interoperability involve standards development. The main factors that affect geoscience interoperability standards development are founded on social, geoscientific, and technological concerns (Fig. 27.4).

Given the age and maturity of geology as a science, one may have thought that numerous international (and national) standards published by organizations such as ISO, CEN, DIN, or BSI would be available. However, apart from the global standards of the International Union of Geological Sciences (IUGS) on rock age and lithology, there are not many worldwide, certified, official sets of geological, topographical, or cartographical standards for geological information, and mainly the INSPIRE standards [5, 12]. Also, there are a set of ISO standards for graphic symbols representing sedimentary, magmatic, metamorphic rocks and tectonics on maps [13–19] (ISO 710-1 to ISO 710-7, 1974–1989). These standards are defined for graphic symbols used on detailed maps, plans, and geological cross-sections.

In addition, EN ISO 14688 *Geotechnical Investigation and Testing—Identification and Classification of Soil* [20] tackles the classification of sedimentary material, including grain size, while EN ISO 14689-1 *Geotechnical Investigation and Testing—Identification and Classification of Rock* [21] deals with the description of rock for engineering geology investigations. Both of these are important sources for rock classifications within ontologies such as the IUGS Commission on the Management and Application of Geoscience Information (CGI) vocabulary and GeoSciML ontology (see later).

The International Stratigraphic Chart (ISC) of the International Stratigraphic Commission of the IUGS provides a thoroughly peer reviewed and globally accepted standard that defines the age of a geological unit (Fig. 27.5). The colors used in the ISC are also used to portray the ages of geological units within the Portrayal Rules of Technical Guidelines of the EC INSPIRE Data Specification for Geology [22]. Additional standard colors and styles are used to describe, e.g., lithology and fault types within the geology theme (see the section on INSPIRE later in this chapter).

The ISO 19100 series standards provide the basis for the interoperability of geoscience GIS, in particular distributed data sources. Geoscience Markup Language [24] (GeoSciML) was developed under the umbrella of the IUGS Commission on the Management and Application of Geoscience Information [25] (CGI), and provides a good example of the application of the ISO 19100 series within the geological community to make geological information interoperable and provide the basis for harmonization (see later). In 2017, GeoSciML (v 4.1) was ratified and published by the OGC. A joint OGC/CGI GeoSciML Standards Working Group (SWG) is continuing the development of GeoSciML.

GeoSciML

GeoSciML is a GML application schema that is designed to transfer information about geology. It provides a framework for the application-neutral encoding of thematic geoscience data and related spatial data and was created to support the interoperability of systems and data. GeoSciML is

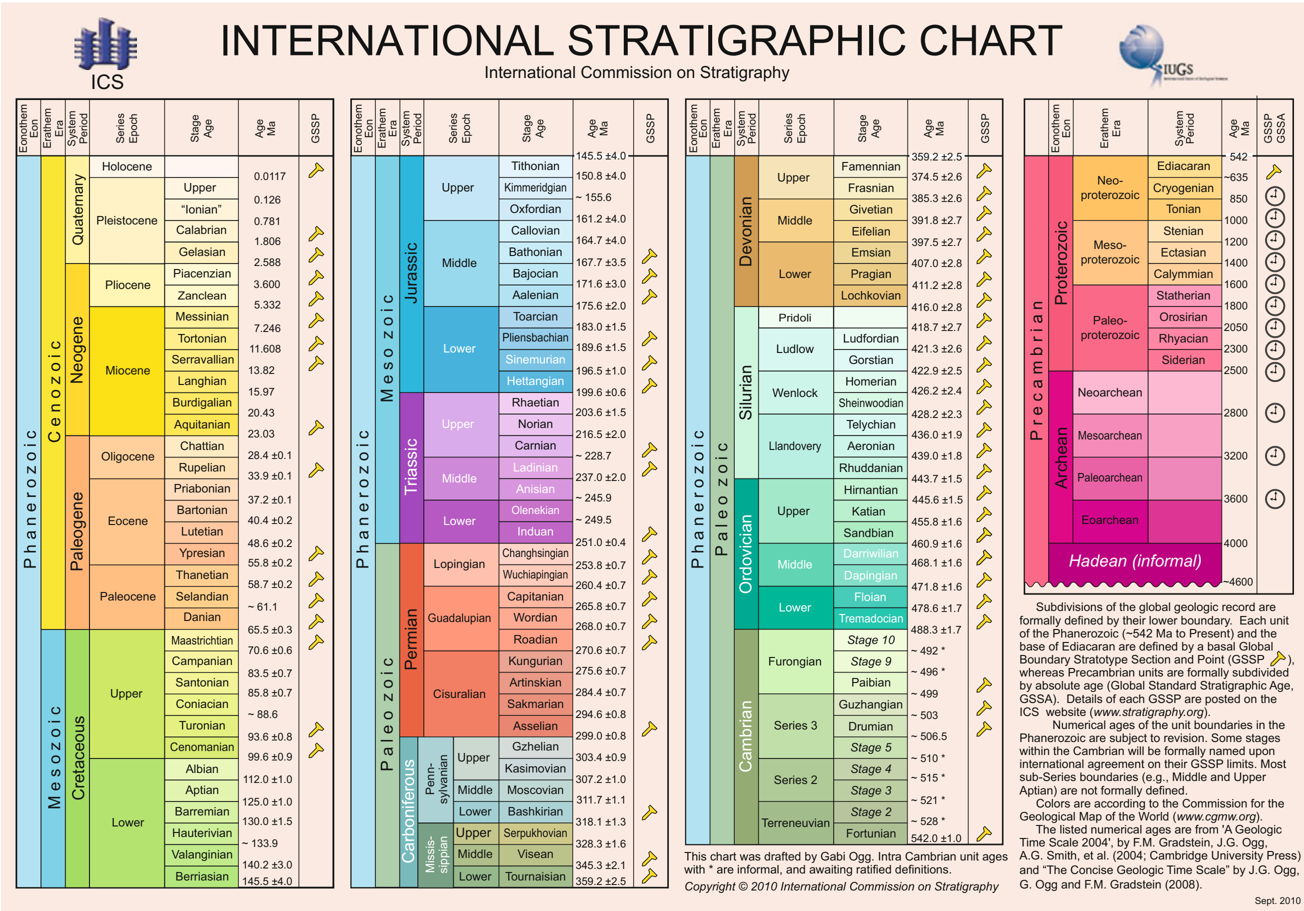


Fig. 27.5 International Stratigraphic Chart of the IUGS [23]. Reproduced with permission, © ICS International Commission on Stratigraphy 2019

based on several ISO and OGC standards: ISO 19136-1 *Geography Markup Language* (GML) [26] for the representation of features and geometry, ISO 19142 *Web Feature Service* (WFS) [27], and ISO 19156 *Observations and Measurements* [28] for observational data. Geoscience-specific aspects of the schema are based on a conceptual model for geoscientific concepts and include geological unit, geological structure, and Earth material from the conceptual North American geologic map data model [29] and borehole information from Exploration and Mining Markup Language (XMML). Controlled vocabulary resources for specifying content to realize semantic data interoperability were developed and tested in particular by the OneGeology-Europe project.

Having agreed on a conceptual data model, it needs to be mapped onto an interchange format. The GeoSciML application is a standards-based data format that is intended for use by data portals publishing data for customers in GeoSciML, when interchanging data between organizations that use different database implementations and software/systems environments, and, in particular, in geoscience web services. Thus, GeoSciML allows applications to utilize globally distributed geoscience data and information. This has been realized within several projects such as the global OneGeology project portal, the OneGeology-Europe portal, the US Geoscience Information Network [30] (USGIN), the INSPIRE geology theme, and Australian projects such as AuScope [31] and AUSGIN [32]. Also, the EC Directive INSPIRE uses it for its Geology and Mineral Resources Implementing Rules, and has adapted the GeoSciML and Earth Resource ML (ERML, see later) [33] data models and CGI vocabularies.

GeoSciML is *not* a database structure. GeoSciML defines a format for data interchange. Agencies can provide a GeoSciML *interface* on their existing database systems, with no restructuring of internal databases required (Fig. 27.6). GeoSciML includes a logical model and GML/XML encoding rules for geological map data, geological timescales, boreholes, and metadata for laboratory analyses.

Scope of GeoSciML

The scope of GeoSciML is mostly interpreted information shown on geological maps, but it also includes observational data from boreholes and field observations using the OGC observations and measurements (O&M) specification (ISO

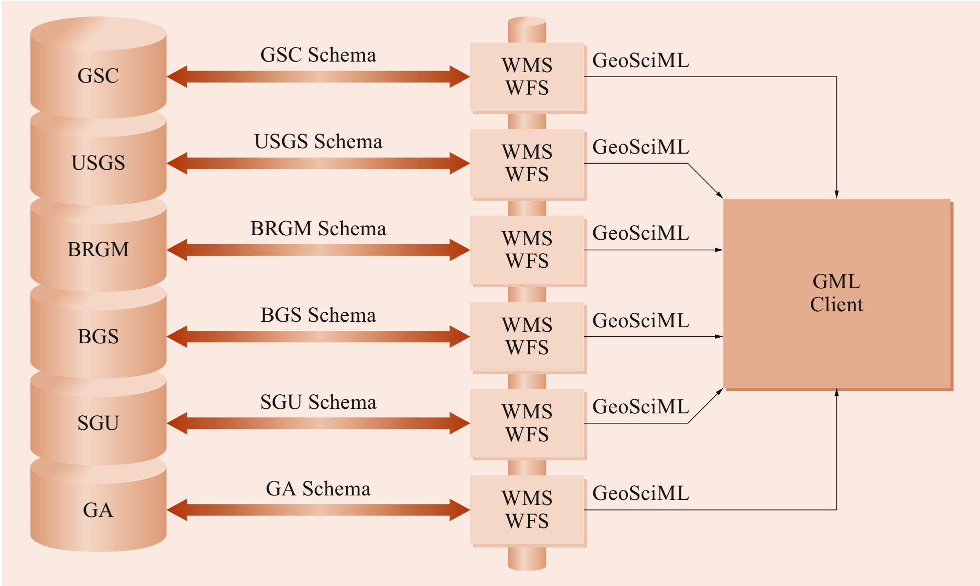


Fig. 27.6 Architecture of the GeoSciML Test Bed 2. GSC – Geological Survey of Canada, USGS – United States Geological Survey, BRGM – Bureau de Recherches Géologiques et Minières (France), BGS – British Geological Survey, SGU – Sveriges Geologiska Undersökning (Sweden), GA – Geoscience Australia, WMS – Web Map Server interface, WFS – Web Feature Service. Figure reproduced with the kind permission of Boyan Brodaric, Natural Resources Canada, and the Commission for the Management and Application of Geoscience Information of the IUGS [34]

19156). The GeoSciML model does not provide definitions for the whole scope of geoscience because other groups may have governance of them.

GroundwaterML is an example of a derived implementation of GeoSciML. It is also the first official collaboration between GeoSciML and an external exchange model group.

EarthResourceML (ERML) is also developed by the CGI, and is an XML-based data transfer standard for the exchange of digital information for mineral occurrences, mines, and mining activity. EarthResourceML describes the geological characteristics and settings of mineral occurrences, the commodities they contain, and their mineral resource and reserve endowment. It is also able to describe mines and mining activities and the production of concentrates, refined products, and waste materials. EarthResourceML makes use of the existing GeoSciML data standard to describe geological materials associated with mineral deposits. It is also underpinned by established OGC and ISO standards, including the Web Feature Service (WFS, ISO 19142), Geography Markup Language (GML, ISO 19136-1), and SWE Common (Sensor Web Enablement) [35]. ERML was used in European projects such as EURARE [36] and Minerals4EU [37].

In 2017, GeoSciML was approved as an OGC standard. This was a significant milestone that established this markup language as an international geological data transfer standard. In collaboration with the OGC, the CGI is continuing to review GeoSciML. GeoSciML is now hosted on the OGC GeoSciML website, but the original CGI GeoSciML website still permits access to all historic versions of GeoSciML [38]. Also, OGC services can be deployed for sensor web applications.

27.3 Maps and Models

27.3.1 Geological Maps

Analog and digital geological maps provide base data that may be utilized in many different ways. The following is a list of important applications of geological maps:

- As a source of information for scientists or scientific institutions to perform further research
- For groundwater exploration (which requires knowledge of the stratum)
- For exploration: mineral commodities such as gravel, lime, and metal are bound to certain geological formations or events such as volcanoes or granite intrusions
- For environmental protection: geological data play an increasing role in environmental protection, such as in the mitigation of landslides, site planning, or spatial planning
- For regional planning and infrastructure, e. g., the construction of bridges, tunnels, and buildings
- For superregional, countrywide, and Europe-wide investigations of specific geoscientific topics such as Cretaceous limestone or regions endangered by ground failure.

In view of these manifold applications, the creation of interoperable, consistent spatial geological databases is of the highest priority. Modern database systems need to not only guarantee the output of informative, clear, and easy-to-understand products (such as map products) but also to allow the creation of thematic maps according to very specific queries.

27.4 3-D Geological Models

Geologists, along with many other scientists, have always been spatial thinkers with the ability to visualize 3-D geological formations and structures in the subsurface. Maps presented as plane views, however, are limited to showing the distribution of geological units at the surface or at defined buried levels, e. g., the top of the bedrock or unconformities. Variations in the depths and thicknesses of 3-D surfaces and volumes have traditionally been depicted in plane view through the use of contours, isopachs, or color shading. Historically, it soon became clear that mine planning and engineering required some form of 3-D mapping, and the two principal approaches to 3-D geological models—modeling mine volumes as either polyhedra or voxels—were developed in the late nineteenth century (Fig. 27.7; [39, 40]) and used in analog models (Fig. 27.8).

In the context of mapping operations with more data and spatial controls, such as mine planning, 3-D modeling software is well established. Such software enables accurate visualizations of diverse types of spatial information that are much too complex for unaided human comprehension. These visualizations also lead to a common understanding of, or a debate about, geological structure and subsurface conditions and the production of the 3-D geological framework model that can be defined as the geological map extended into 3-D. These block models can be viewed from any angle at a range of resolutions, taken apart and reassembled, sliced (cut in one direction), or diced (cut in multiple intersecting directions) for further visualization, checking, quality assurance, interrogation, and analysis. So, once constructed, a geological framework model or block volume can be used to produce automated outputs such as borehole prognoses, lines of vertical section, uncovered surfaces, horizontal slices

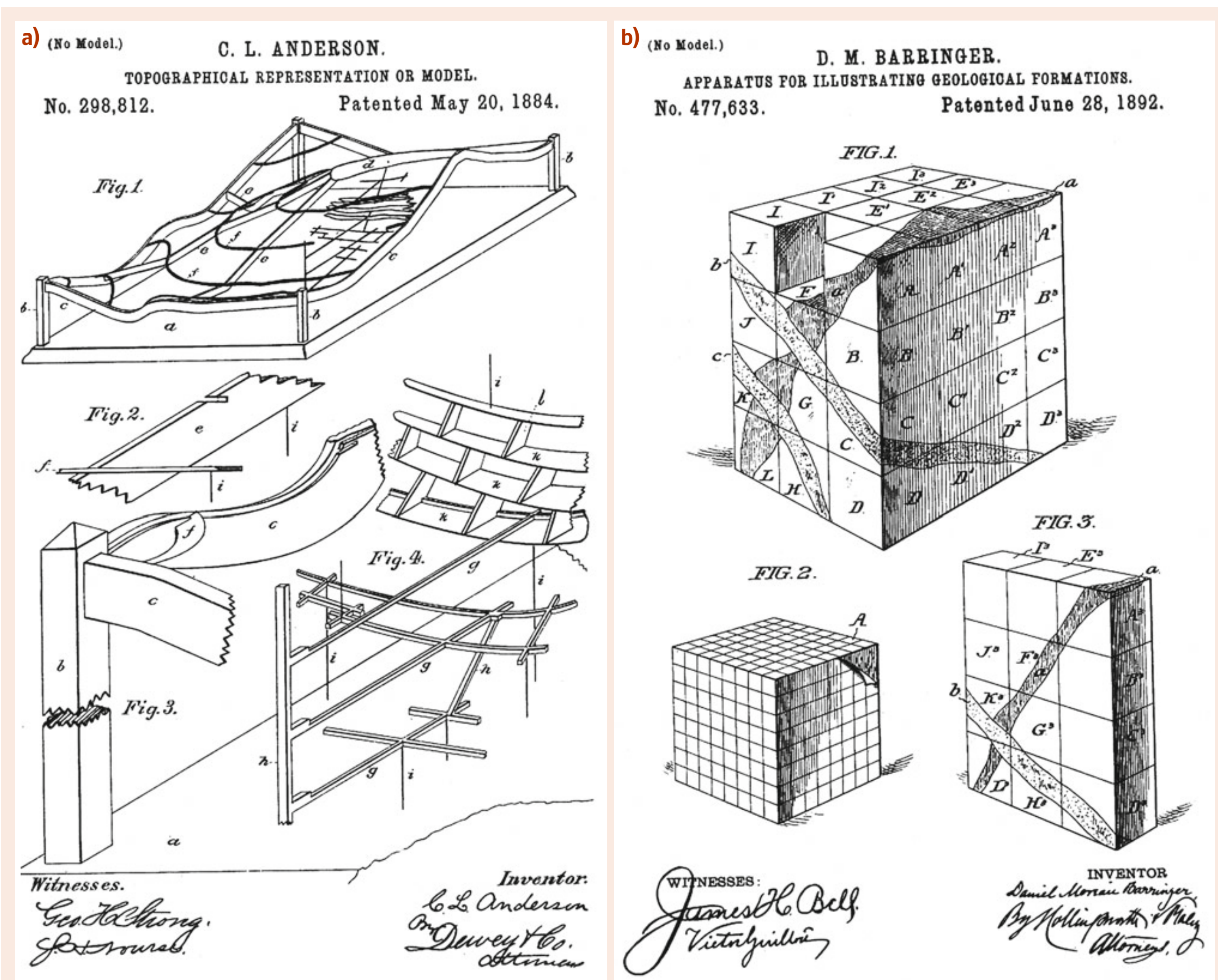


Fig. 27.7 Early examples of 3-D models used in mine planning: (**a**) inset Fig. 4 gives an example of mine levels, shafts, and tunnels depicted in three dimensions [39]; (**b**) the geological structure is structured into voxels that can be taken apart for visual analysis [40]

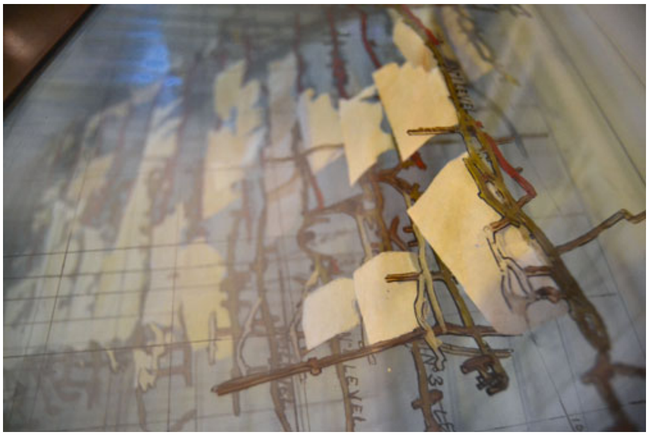


Fig. 27.8 Detail of a 3-D model of the Sons of Gwalia Mine (1897–1963), Gwalia, Western Australia. Photo by J. Klump

of the geology at defined elevations or specified depths from the surface (reduced DTM), and volumetric calculations.

The construction of 3-D models of the geology forces the geologist to populate the whole spatial volume, making assumptions about the geometry of the subground, while conventional geological maps may leave the details up to interpretation. The greater rigor demanded by 3-D modeling generates useful debate and discussion amongst teams, and it is hoped that it will lead to improved concepts, interpretations, and the formulation of new research ideas. In this way, the current ideas and concepts of the geologist are captured, and the uncertainty in an interpretation can be better described.

With graphic computing becoming available, the combination of GIS and computer-aided design (CAD) has enabled the portrayal of geological units as a series of individual surfaces that can be visualized in three dimensions. Three-dimensional geological modeling has developed dramatically over the past 30 years from contouring and gridding techniques using mainframe computers through to PC-based geological modeling software pioneered mainly by the hydrocarbon and mining industry. The relative simplicity of hydrocarbon deposits compared to mineral deposits and the large exploration budgets available in the hydrocarbon industry have allowed great advances in the field of 3-D modeling of hydrocarbon deposits. On the other hand, the complex geology and multivariate geochemistry of mineral deposits pose additional challenges that have caused the 3-D modeling of mineral deposits to trail behind.

In general, the tools for 3-D geological models have become customized to deal with very specific user requirements, geological scenarios, and data types. As a result, these systems are often expensive to license, complex, and require specialized modelers to operate them. Furthermore, software vendors have shown little interest in facilitating interoperability between different product lines, resulting in *vendor lock-in*, where it has become next to impossible to transfer models from one software to another. CAD and GIS tools have also been developed to deal with geological environments, but these have often led to convoluted multisoftware solutions that become hard to use and implement as a single simple workflow.

To overcome these barriers, the Loop consortium [41] was formed as a OneGeology initiative. The Loop consortium brings together geological surveys and research institutions in Australia, Canada, France, Germany, and the UK with the goal of founding a new open source initiative to build the next generation of 3-D geological modeling tools. The project aims to integrate all the available data in the modeling process and develop enabling technologies that will combine probabilistic modeling with structural concepts to produce 3-D geological models. The models will allow uncertainty to be assessed and characterized throughout the modeling workflow in order to optimize further data acquisition and thus optimally reduce uncertainty.

Many geological survey organizations (GSOs) such as the British Geological Survey (BGS) have implemented software systems and methodologies to facilitate the complete migration from 2-D paper-based to 3-D digital geoscience information, such as geological framework models [42]. As with geological maps, these models are used in research but also as an aid when planning and managing the subground. Common users are government departments and regulatory agencies as well as the resources industry. In addition, 3-D models are a great resource for geoscience education at all levels, to guide and inform decision makers, and also to show the public what goes on under our feet and why decisions are often made based on this understanding.

27.5 Types of Geological Models

Geological modeling packages, and to some extent the modelers, employ one of two different approaches to modeling: they are either expert controlled (also known as explicit), or data driven (also known as implicit, stochastic, or numerical). Both approaches have their advantages and drawbacks, which depend on the background data, the geological setting, and the user requirement for the particular model.

27.5.1 Expert-Controlled Geological Models

Expert-controlled modeling should utilize all the available hard data but then deliberately insert sufficient expert-controlled soft data to populate the model to calculate a geological model that makes sense and takes into account our current understanding of the geological system. These data might, for example, involve drawing a network of cross-sections that are included in the model calculation. Embedded in these sections are the explicit shapes of the contacts between units, which may be derived based on the

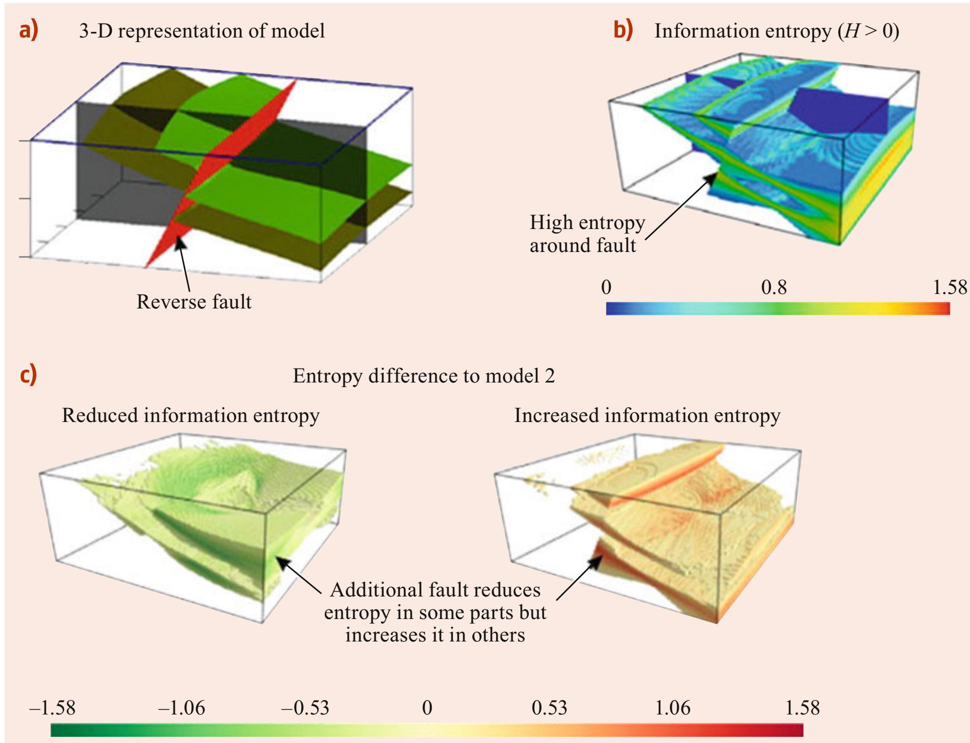


Fig. 27.9 Results of a geological hypothesis test, model 3: (**a**) 3-D representation of one possible realization with a reverse fault offsetting the geological units; (**b**) information entropy ($H > 0$) shows that uncertainties are particularly high around the fault; (**c**) upon comparing model 3 to model 2 with difference plots, we observe that uncertainties are reduced in some parts of the model (*green colors*) but are greatly increased in other parts (*red colors*). Reprinted from [43] with permission from Elsevier

rules of stratigraphy and experience of seeing similar rocks at an outcrop or by distinguishing between differing styles of bed arrangements—for example, onlap, offlap, overstep, erosional, or channeled relationships. It should always be remembered that geology is essentially an interpretive science, as opposed to say (geo)chemistry or (geo)physics. The main advantage of expert-controlled models is the ability to get geologically sensible results at the first attempt, drawing on the holistic knowledge of the most experienced geologist(s) available. Conversely, the results are not reproducible and the uncertainty is harder to quantify unless it can be derived solely from the hard data distribution.

27.5.2 Implicit Geological Models

Data-driven models are those in which observed and measured data (including geological interpretations) are treated as the entire valid dataset. Calculation is purely by mathematical or geostatistical interpolation and by extrapolation from known data points.

This approach is advantageous in that it is totally objective, totally reproducible, and very suitable for numerical data, so it is easy to quantify uncertainty mathematically. An example might be the zonation of grades in a buried ore body from analytical measurements of regular spaced core samples. On the downside, whilst they obey the rules of mathematics and statistics, some calculated models may defy the laws of geology, or—more commonly—omit the geologists' knowledge and understanding. One solution that is often used when the results of a model calculation fail to exhibit geological common sense is to constrain the hard data by inserting phantom borehole data or by forcing the model to fit an interpreted cross-section. This involves the introduction of soft (interpreted) data, as in expert-controlled modeling, resulting in a mixed approach. Further, it should be noted that some of the supposed hard data used in data-driven models is not always as hard as one might wish to believe. For example, geologists frequently disagree about the positions of stratigraphic boundaries (picks) in boreholes, at outcrops, and in seismic profiles.

A shortcoming of many common 3-D modeling systems is that they do not facilitate a rapid revision and iteration cycle such as that required in an active national or regional geological survey [44, 45]. This rigidity is a consequence of the CAD approach to building the model, in which all elements of the model are explicitly defined. The high cost of creating variations of a model also discourages the use of ensemble modeling to explore the uncertainty in the model. A different approach to geological mapping and the construction of geological 3-D models is to use implicit geological models [46]. An implicit geological model only takes the data points available from observations and then uses them as anchor points to extrapolate the geometries of geological units [47]. In addition, the resulting model can be used as the basis for a forward model of the resulting gravity and magnetic field. The resulting geophysical model data can then be compared with geophysical observations in the field. A correct model must satisfy both the geometric and physical ground truth provided by the field observations [47].

The concept of implicit geological models takes into account the inherent uncertainty in our knowledge of the exact composition and geometry of the subground by relying purely on observations. The presence and detection of uncertainty is an important issue [48], and has recently received significant attention [43, 46, 47]. Implicit models can be used to explore the inherent uncertainty in a model through ensemble modeling. In this approach, the input parameters of the model are varied systematically. The resulting models can then be compared as an ensemble and the localized uncertainty expressed as information entropy ([43], Fig. 27.9).

27.6 Data Structures

In two dimensions, a geologic formation or unit is represented by a polygon, which can be bounded by faults such as lines or unconformities or by its lateral extent or crop. In geological models, a geological unit is bounded by 3-D triangulated or gridded surfaces that can be composed of a combination of faults, tops, or bases of the units. The 3-D equivalent to the mapped polygon is the fully enclosed geological unit presented as a triangulated mesh; this is often referred to as a shell (or an object or volume). For the purpose of property or fluid modeling, these volumes can be separated further into an array of 3-D cells, often referred to as *voxels* (combining the words *volumetric* and *pixel*). Just like their 2-D equivalent, they discretize space and enable the assignment of properties (such as hydraulic conductivity or confidence values) to each of these cells. Voxel grids can be made up of either orthogonal cellular meshes, quadtree meshes, or unstructured triangular meshes (also called tetrahedral meshes) [49]. These property meshes are needed to make the geological framework models useful and usable in property and process modeling. Figure 27.10 illustrates the differences and commonalities between geological data rendered in 2-D (paper or GIS) and 3-D.

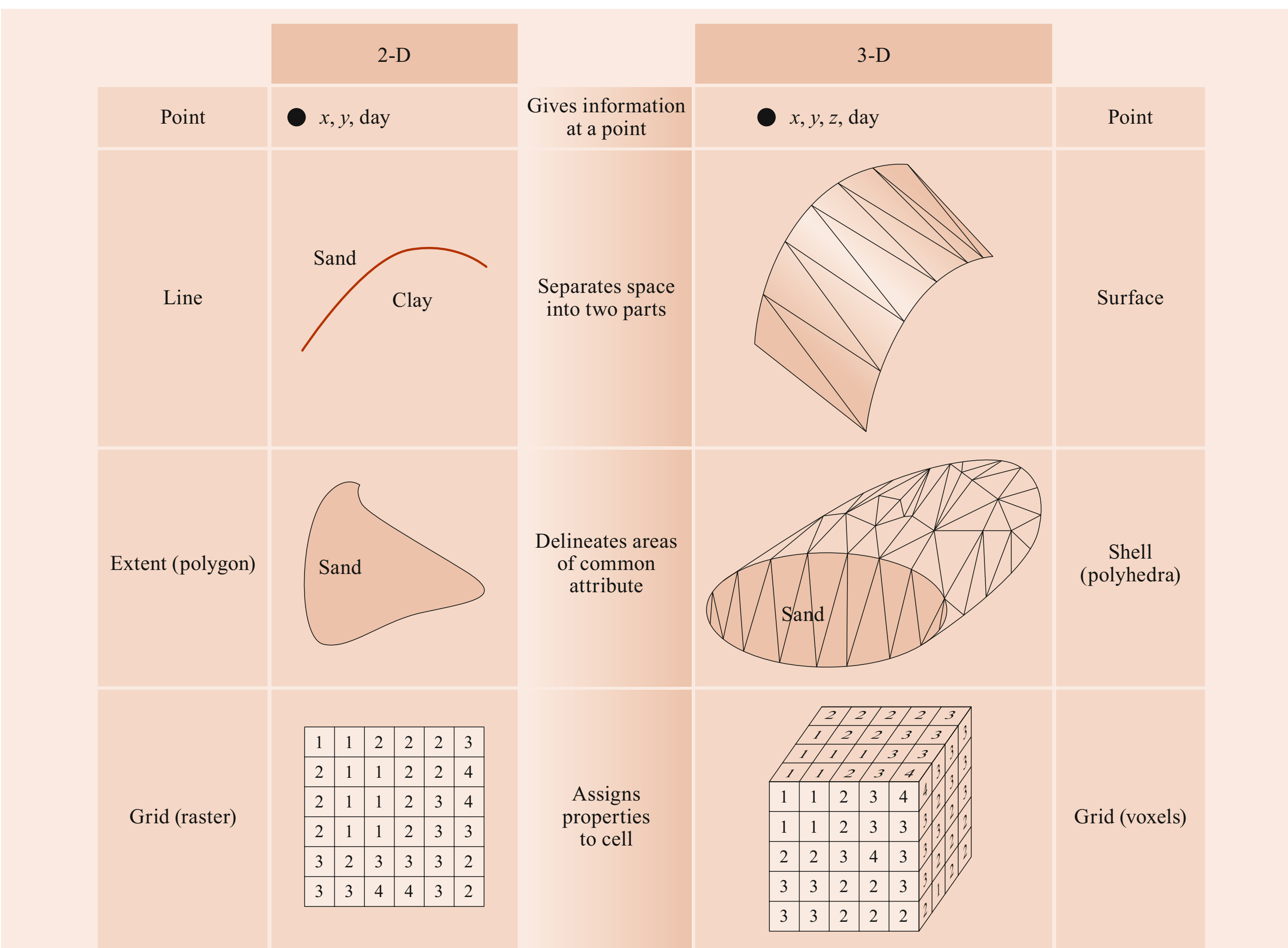


Fig. 27.10 Differences and commonalities between geological data rendered in 2-D (paper or GIS) and 3-D. In two dimensions, a geologic formation or unit is represented by a polygon, which can be bounded by faults such as lines or unconformities, or by its lateral extent or crop. In 3-D geological models, a geological unit is bounded by 3-D triangulated or gridded surfaces, which can be composed of a combination of faults, tops, or bases of the units. The equivalent to the mapped polygon is the fully enclosed geological unit delineated using a triangulated mesh; this is often referred to as a shell

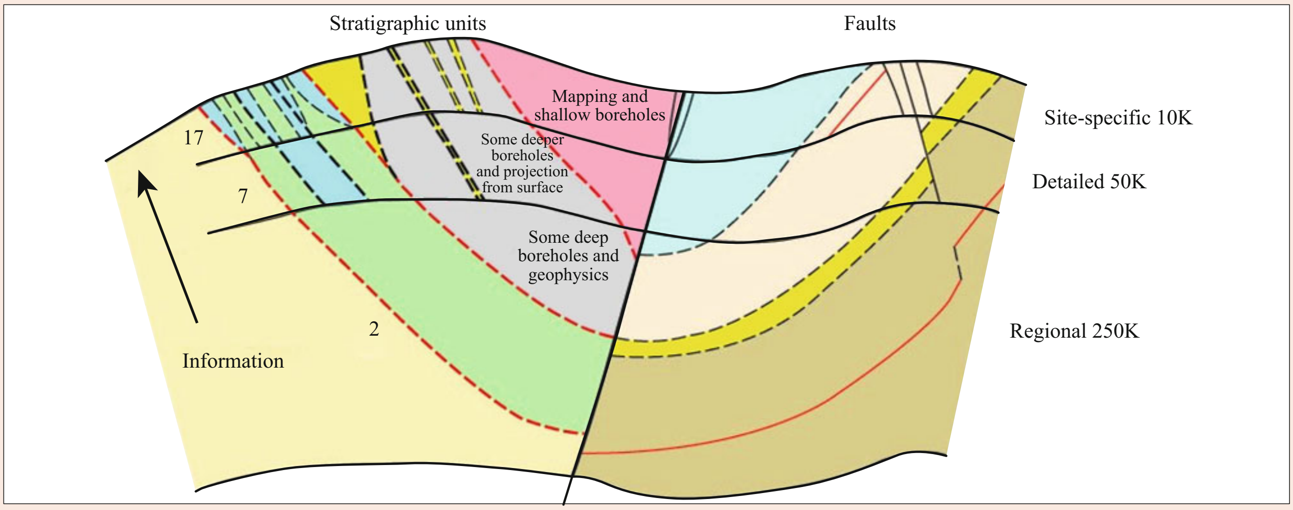


Fig. 27.11 Interrelation between model resolution and depth of investigation

Fig. 27.12 The 1 million resolution onshore model of Great Britain

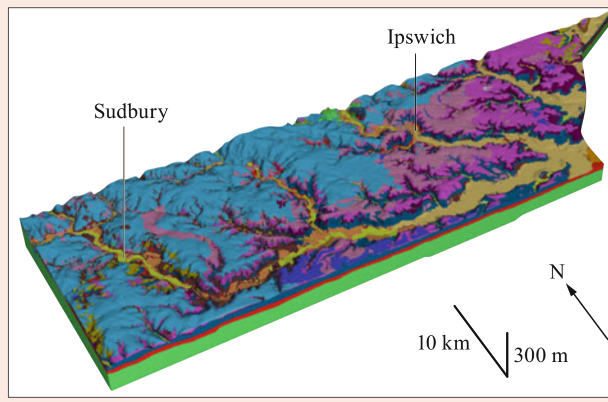


Fig. 27.13 The 50k resolution southern East Anglia model of the Ipswich–Sudbury area, covering 1200 km^2

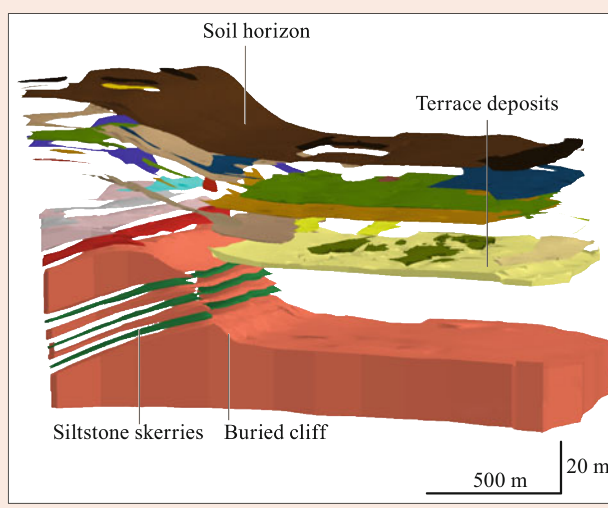


Fig. 27.14 The site-specific Shelford model

An advantage of digital geological models is that they are scale independent. Using the term *resolution* to substitute for the scale of a geological map, a 250k resolution model is one that contains a level of surface geology similar to a 1 : 250,000 scale geological map but remains suitable for use at a range of resolutions centered around 250k resolution. In practice, 3-D framework models tend to be constructed, like their map predecessors, at a range of resolutions required for specific tasks. Often, high-resolution models of a small region are nested in a lower-resolution model of a broader region. Central to this concept is that the varied resolutions are consistent with each other, so that they collectively form a seamless transition from a general national-scale model to a detailed or site-specific one. Conversely, detailed models can be generalized and incorporated into coarser-resolution versions (Figs. 27.11–27.14).

27.7 3-D Geological Modeling Software

Considered collectively, GSOs deploy various software packages. The most prominent and most suitable for the needs of a geological survey appear to be GoCAD, 3D GeoModeller (widely used in GSOs in Australia, North America, and France), and GSI3D, used by BGS. GSI3D [50] is a totally expert-controlled package, whereas GoCAD and 3D GeoModeller follow a mixed data-driven and expert-controlled approach. As discussed above, there are also many other geological modeling systems and methodologies that are extensively documented elsewhere; a full review is beyond the scope of this chapter [45, 49, 51–54].

27.7.1 GoCAD

Developed from a project started in 1989 by Jean-Laurent Mallet at Nancy Université and now marketed by Paradigm, GoCAD has been used by specialized 3-D modelers within BGS and many other GSOs together with the hydrocarbon industry for at least a decade for the construction of 3-D geological models. Methodologically, as its name suggests, GoCAD is a CAD system that requires the input of data and then applies complex, proprietary interpolation and surface-fitting algorithms. GoCAD has been used successfully at GSOs such as BGS and Geoscience Australia for the construction of regional and national resolution models based mainly on deep borehole data and depth-converted seismic interpretations [55].

27.7.2 3D GeoModeller

The other main geological modeling package in use by GSOs is 3D GeoModeller, which developed from a requirement by the French Geological Survey (BRGM) to create a geological editor instead of using CAD or GIS techniques. BRGM believed it was unnatural to force geologists to think in a way that is contrary to their training in order to create a 3-D model. Therefore, a research and development project known as GeoFrance 3-D was set up; this ran for 6 years, leading to the development of the prototype 3DWEG (3D Web Editeur Geologique) tool, the precursor to 3D GeoModeller. More information on this modeling package is available in [56].

27.7.3 Leapfrog

Leapfrog is a 3-D implicit modeling software package that was first introduced to the market in 2003. It has since established itself as a standard tool for 3-D geological mapping in the mining sector. Leapfrog's implicit modeling engine bypasses the construction of wireframes and generates surfaces directly from data by fitting a surface through the data points, thus removing the need for manual digitization of the geometries of geological bodies and allowing for frequent updates as new data are added to the model. The fitting algorithm uses a "fast form" implementation of Radial Basis Functions (RBF) to interpolate between data points in 3-D space [57]. The fast form implementation reduces the complexity and compute time for solving RBF and making them suitable for practical applications.

27.7.4 3-D Geology in Web GIS

GIS are expert tools that require some training before they can be operated by a user. Advances in web browser technology have led to the development of visualization and dissemination tools that make geological models useful for and usable by end users such as planners and regulators and useful for communicating with the public. GSOs and other producers of 3-D models have been innovative in making use of existing technologies such as browser-based 3-D visualizations that use X3D or VRML, Google Earth, and 3D PDF. The advantage of these technologies is that they are available globally at no cost to the end user and are relatively simple and intuitive to use.

In order to deliver the full richness of geological models and give the user full interaction with the models it is necessary to develop entirely new software systems. The BGS LithoFrame Viewer is an example of a standalone product for the delivery and analysis of geoscience models that partly fulfills the above requirements. The functionality of the Subsurface Viewer includes uncovered maps, synthetic boreholes and slices, synthetic sections, views of single geological objects, block models, exploded views, and the ability to switch between different properties of the geological model (Fig. 27.15). A small demonstration model accompanied by a user manual [58] is given at [59].

27.7.5 Environmental and Subsurface Management

Managing the subground—an important task since the early days of industrialization—initially centered on mining and engineering applications. In recent decades, further considerations have been added to this task, such as the management of groundwater resources, nuclear waste disposal, and carbon

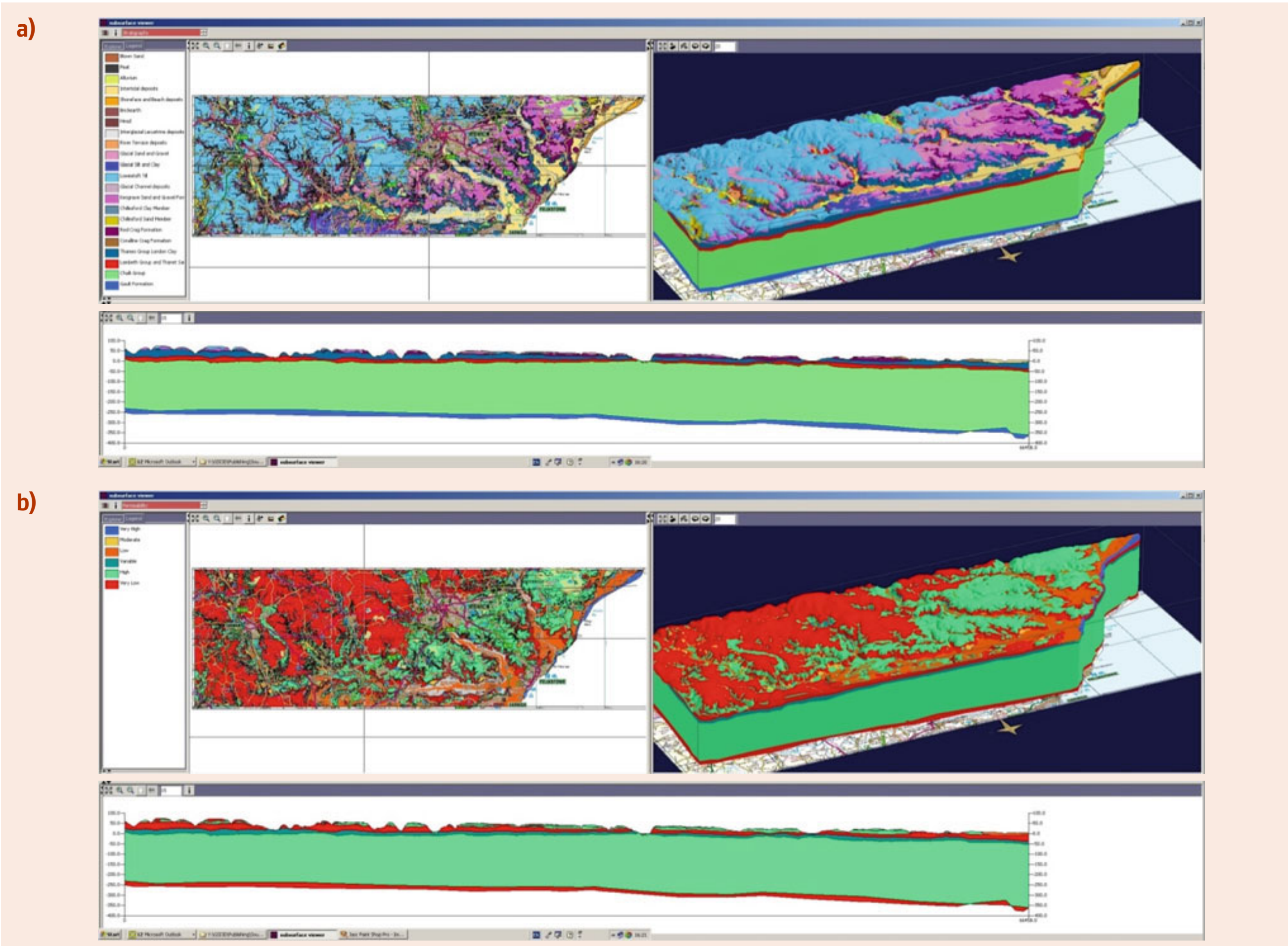


Fig. 27.15 The LithoFrame Viewer interface, showing the southern East Anglia Model with attribution by stratigraphy (**a**) and permeability OS (ordnance survey topography) (**b**)

dioxide sequestration. The effects of natural disasters and the impact of global climatic change at the Earth's surface need to be managed and mitigated.

Reports by *Stern* in 2006 [60] and by the Intergovernmental Panel on Climate Change [61] have identified the need for urgent action to tackle the causes of climatic and environmental change and, in particular, to mitigate their impacts, as well as the need to persuade people to adopt lifestyles that are ecologically sustainable. The Earth sciences will play a major role in monitoring, forecasting, and developing mitigation strategies to address issues such as:

- Providing clean and affordable drinking water for human and industrial use
- Managing and forecasting flooding events
- The disposal, containment, and remining of anthropogenic waste
- The prediction of ground conditions for construction
- Managing and forecasting natural hazards such as landslides, subsidence, earthquakes, and tsunamis

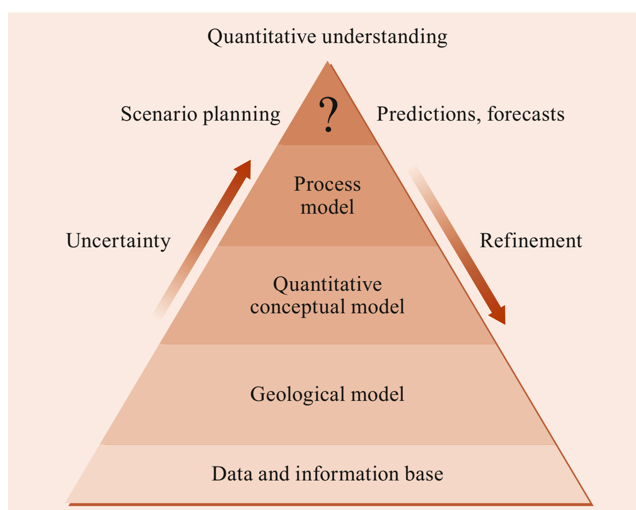


Fig. 27.16 Relationship between geological data, geological 3-D models, and 3-D to 4-D process modeling (forecasting), partly based on [62]

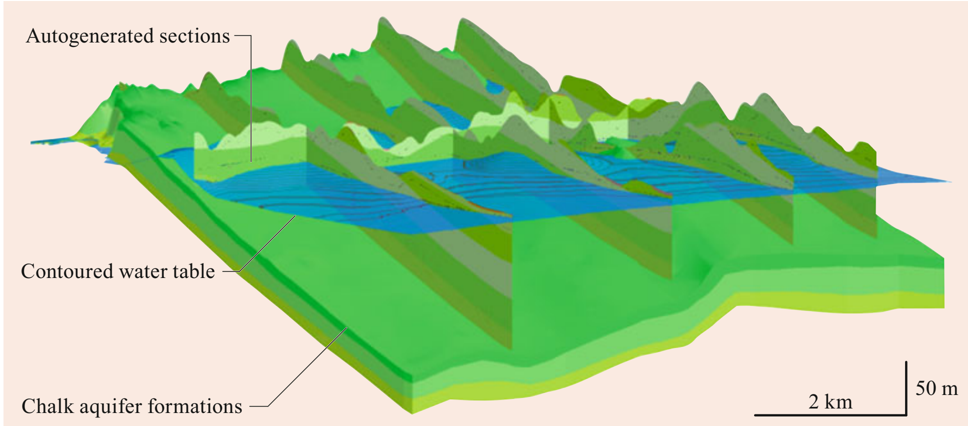


Fig. 27.17 Integrated geological and groundwater models of an area around Brighton, UK (after [63])

- Effective land-use planning and guaranteeing security of supply for food, minerals, and energy while maintaining biodiversity.

It would be impossible for just one scientific discipline in isolation to address these issues or understand the complex interrelationships involved. Many GSOs recognize the need to share data, models, and knowledge with the wider science community to tackle the issues identified above. 3-D geological models form the spatial framework within which surface and subsurface processes are studied. Modeling these processes has become a multidisciplinary activity (geological, conceptual, and process modeling), and the results are used in forecasting to make predictions and projections that enable decision making in environmental management based on a quantitative understanding (Fig. 27.16).

27.7.6 Environmental Modeling Platforms

Several GSOs worldwide already produce 3-D geological framework models that can be integrated with groundwater data (Figs. 27.17, 27.18). Additionally, routines for adding geological constraints to groundwater models are emerging [62, 64], but these efforts are only aimed at solving the groundwater part of the Earth's system.

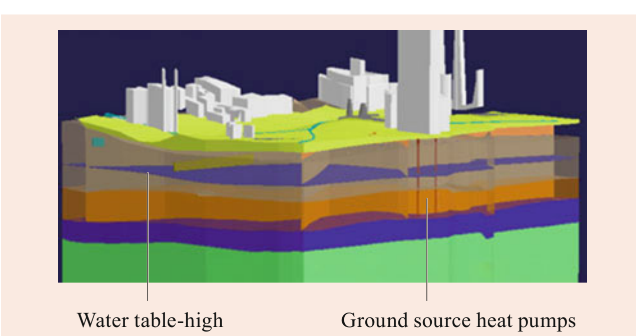


Fig. 27.18 An example of a model of part of central London; artificial infrastructure is integrated into the geological model and visualized using the LithoFrame Viewer

The next logical step is the development of environmental modeling platforms (EMPs) to enable the assembly, inspection, and interpretation of a wide variety of datasets across the environmental sciences. Also implicit in the EMP vision is the need to establish common international data standards and formats to achieve interoperability and data exchange [65, 66].

27.7.7 The Need for Subsurface Management Systems and Regulation

Economic development driven by population growth has placed greater demands on the subsurface, including the need to accommodate transport, utilities, and telecommunications infrastructure. The subsurface is an integral part of the economic system, and the increasingly complex use of the subsystem suggests that a more coordinated approach to its regulation and management is essential and should be coupled with that of the surface. To provide a basis for the sustainable management of the subsurface, a 3-D and 4-D understanding of the ground, ground installations, land-use history, and the suitability of the ground for future use is critical [67].

The key to the delivery of the results from EMPs to planners, regulators, and other decision makers is to make the results visible in the context of the real world, including the subsurface. There is an identifiable need for a comprehensive 4-D subsurface management system that provides the basis for spatial, volumetric, and temporal decision making in the subsurface in the same way as today's GIS are used for spatial planning, insurance risk assessment, or emergency planning.

Systems that fulfill parts of this strategy are emerging at several GSOs. An example from the BGS is given in Fig. 27.18. It shows the integrated visualization of anthropogenic structures such as tunnels and buildings within the context of a geological model.

27.8 Spatial Data Infrastructures

A spatial data infrastructure is a coordinated series of agreements on technology standards, institutional arrangements, and policies that enable the discovery and facilitate the availability of and access to spatial data. SDIs are mainly built by governments (regional, national, international). Examples include INSPIRE (the European Commission initiative to build a European SDI beyond national boundaries), the United Nations Spatial Data Infrastructure (UNSDI), the United States National Spatial Data Infrastructure (NSDI), and EMODnet—the European marine SDI that includes thematic data on a range of subjects such as geology, bathymetry, and seabed habitats.

27.8.1 INSPIRE

For several decades, a treasure trove of spatial data has been collected across Europe by governmental organizations. However, this information is commonly hidden away and difficult for the public to access. Moreover, it is stored in a variety of data formats (some still on paper) according to different scientific and technical classifications and various naming standards based on different reference systems, and, inevitably, in different languages. This leads to situations in which urgently needed information is not only difficult to find but also barely usable, even by experts. Reworking and adapting these data, assuming they are finally found, is a long and costly process that only rarely leads to a satisfactory result, as the standards required to unify and harmonize the data rarely exist. The consequence is an unnecessary loss to society at large of valuable information—particularly transborder information—for environmental and other applications [67].

With Directive 2007/2/EC, the European Commission is determined to change this situation by creating an Infrastructure for Spatial Information in the European Community: INSPIRE. Its general aim is to provide a legislative framework that enhances the accessibility of environmentally relevant thematic data to politicians, economists, scientists, and citizens in EU member states. Optimally, these data must be consistent and comparable, and they must be provided in their best-validated and most useful form. Integrating the environmental spatial data from all EC countries and establishing their interoperability across political boundaries are also essential aims of the EC Directive and the corresponding acts of EC Member States.

INSPIRE addresses not only the data and their availability and interoperability but also their respective metadata. Within the INSPIRE framework, network services are created that facilitate access to the data and their visualization, to agreements on common data use and access, to test suites for data and service validation, and to applications and mechanisms for the coordination and monitoring of the whole process.

INSPIRE includes 34 thematic (geo)data fields and their metadata in three annexes; these data must be prepared according to the INSPIRE Implementing Rules in a consistent and comparable way.

Five of the 34 themes relate to geoscience information:

- Geology, including lithology, aquifers, and geomorphology
- Soil
- Natural risk zones
- Energy resources
- Mineral resources.

Geoscience data are thus defined *by law* to be environmental data, and rightly so, as they give us essential information about the Earth upon which we live:

- They are fundamental to natural hazard (landslides, earthquakes, volcanoes, and flooding) prediction and mitigation
- They relate to resources and the security of energy supplies (oil, gas, coal, and geothermal resources).
- They are used to calculate the effects of soil and water pollution
- They are fundamental to dealing with the effects of climate change and mitigation (coastal erosion, CO_2 sequestration).

The INSPIRE Directive, adopted through co-decision procedure in March 2007 by the European Parliament jointly with the Council of the European Union, was put into force in May 2007 [68]. The Directive had been transferred into national law by all the European Union Member States by 15 May 2009. Among its other effects, this resulted in new tasks and considerable effort for European geological survey organizations.

Following the principles of equality in a community that currently embraces 27 nations, realizing Directive 2007/2/EC necessitated the formal involvement of committees comprising representatives from Member States, spatial data interest communities, legally mandated organizations, and, perhaps less formally, the involvement of numerous stakeholder parties, organizations, task forces, working groups, etc., in a procedure that is now generally referred to as the INSPIRE Process.

As specified by Paragraph (5) of the INSPIRE Directive [68], INSPIRE is based on the infrastructures for spatial information created by the Member States, which are made compatible with each other through common Implementing Rules (IR) and are supplemented with measures at the community level. These measures ensure that the infrastructures for spatial information created by the Member States are compatible and usable.

Five expert Drafting Teams (DT) on metadata, network services, data specifications, data and service sharing, monitoring, and reporting were established. These DT created the IR that define the terms and requirements for the realization of the INSPIRE directive in detail. INSPIRE exclusively addresses datasets that are the property of public authorities and only targets preexisting data. However, these need to be adapted according to the INSPIRE Implementing Rules [69].

The implementation of INSPIRE encompasses:

- Creating the metadata for all the spatial data by 2010/2013, and then keeping that metadata up to date
- Making thematic spatial data conformal
- Making the data interoperable (by definition, classification, attributes, thesauri, and relationships among the spatial objects)
- Making discovery and viewing services operable by 2011
- Making download services operable by 2012
- Making the interoperable data available by 2017/2020, to help build and feed a network of services that can discover, share, view, and download the spatial data
- Monitoring and reporting the progress of the INSPIRE implementation.

Key dates for the geological surveys were:

2010 The first metadata to be made available.
2012 Data types, attributes and level of detail to be defined for geoscience data.
2015 Newly collected geoscience data to be provided.
2020 Archive data to be provided.

Many geological survey organizations initially tended to see INSPIRE as an imposition and a threat—and some might still do so [70]. Nevertheless, the organizations rose to the challenge and took the opportunity to be actively involved in the INSPIRE process.

- EuroGeoSurveys, the Association of the Geological Surveys of the European Union, is the SDIC (Spatial Data Interest Community) for geological surveys in Europe. It initiated an INSPIRE working group, now called the Spatial Information Expert Group (SIEG), in which European geological survey representatives discuss geoinformation and INSPIRE-related issues.
- Several survey organizations (e. g., French, German, and British) are registered as INSPIRE LMOs (Legally Mandated Organizations), which has benefits, e. g., they have the right to review the IR.
- Several geological surveys (e. g., those of Germany, UK, France, and Norway) have been actively working on the INSPIRE IR, allocating experts to the drafting teams in order to provide active support with scientific input and perhaps even influence the conditions of the implementation of the Directive.
- Twelve geological surveys sent representatives to the INSPIRE Thematic Working Group on Geology and Mineral Resources, which developed the technical and scientific data specification rules for these themes.

This new EU Directive represents an immense opportunity that will raise the awareness of the EU administration, politicians, scientists, economists, and the general public of the importance of geoscience data. In addition, with input from the geological and other communities, INSPIRE is developing and introducing geoscience standards agreed across Europe that might influence or even become globally acceptable standards. Additionally, this Directive is adding value to (often hidden) European survey data by making them internationally standardized—which could even become a new quality level (*our data are INSPIRE-compliant*)—and thus comparable and usable. This will be an invaluable help to cross-border projects. Finally, it is making geoscience data much more discoverable, usable, and laterally comparable and combinable, and thus better known and more widely used.

27.8.2 OneGeology and OneGeology-Europe

OneGeology is a voluntary global initiative to improve the accessibility of geological map data, the interoperability of that data, and the transfer and exchange of know-how and experience [71]. The project aims to create worldwide, dynamic, digital geological and geoscience data web services and increase the accessibility of existing national geological maps and geoscience data services supporting a wide range of digital formats. The target scale is at least 1 : 1 million, but the project is being pragmatic and accepts a range of scales to integrate the best available data. The transfer of know-how is another important approach that recognizes that different nations have differing abilities for participation. The initiative is truly multilateral and multinational and is carried out under the umbrella of several global organizations.

Since its inception in 2006 it has been hugely successful, and 189 organizations from 121 countries are now participating, with over 70 of those nations serving their data to a dynamic web map portal (Fig. 27.19).

OneGeology used the stimulus provided by the UN International Year of Planet Earth 2008 to begin the creation of an interoperable digital geological dataset of the planet. The intent was to design and initiate a multilateral and multinational project that mobilized geological surveys to act as the drivers and sustainable data providers of this

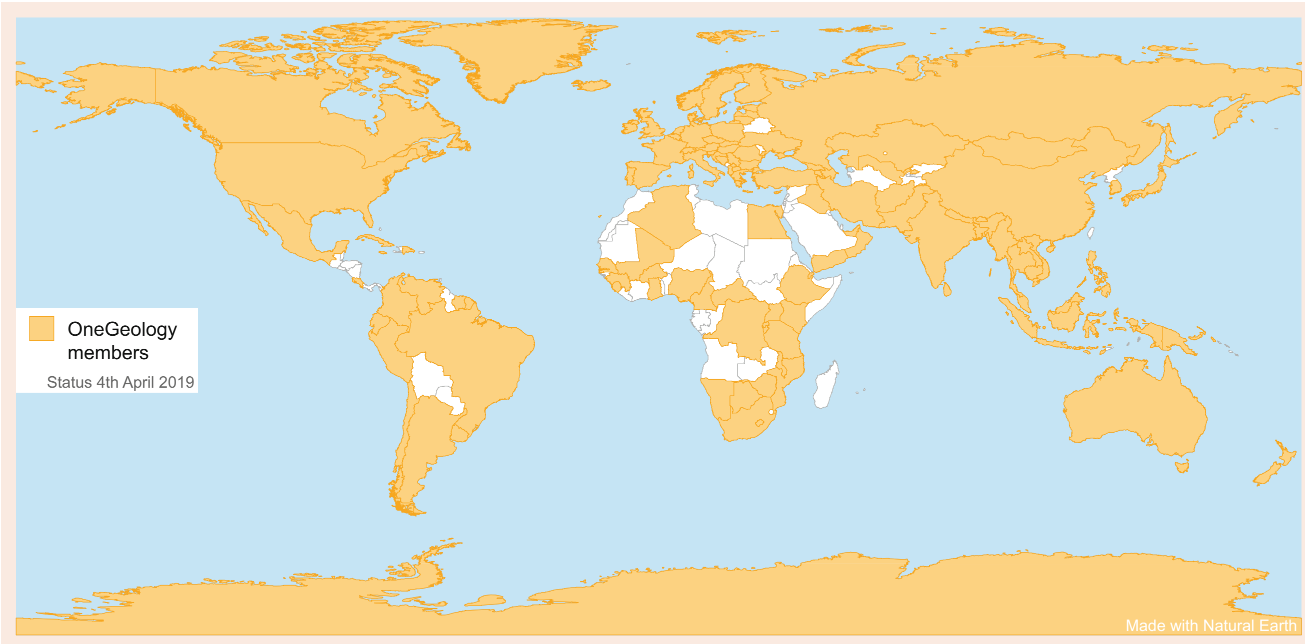


Fig. 27.19 Countries participating in OneGeology who serve national Web Map Servers (WMS) [71]. Reproduced with the permission of the OneGeology project.

global dataset. Now, OneGeology is basically a distributed multinational spatial data infrastructure built on simple operational and technical (WMS and WFS) protocols. It uses the global IUGS-CGI geoscience data model and the interchange standard GeoSciML as a vehicle for creating global geological maps and providing a range of geoscience data. Last but not least, the project is transferring know-how to developing countries and reducing the length and expense of their learning curve while allowing them to serve maps and data that will attract interest and investment. OneGeology is coordinated by the international OneGeology Board and administered by the British Geological Survey. The data portal and metadata services are provided by the BRGM.

27.8.3 OneGeology-Europe

The OneGeology-Europe project made geological spatial data held by the geological surveys of Europe more easily discoverable and accessible via the Internet. The geological survey organization of each EU Member and those of neighboring countries have considerable assets in the form of spatial geological datasets, but before the start of One Geology in 2007, OneGeology-Europe in 2008, and EMODnet in 2009, those assets were difficult to discover and were not interoperable. For users outside of geological surveys, those datasets were not easy to obtain, understand, or use. However, geological spatial data are essential for, e. g., the prediction and mitigation of landslides, subsidence, earthquakes, flooding, and pollution. These issues are global in nature and their profiles have also been raised by the OneGeology global initiative for the International Year of Planet Earth 2008.

Geology is a key dataset in the EC INSPIRE Directive (Annex II). It is also fundamental to several Annex III themes: natural risk zones, energy resources, and mineral resources [72] (see above).

Twenty EU countries participated in the OneGeology-Europe project (1G-E), the main aim of which was to create a web-accessible, semantically and technically interoperable geological dataset at the 1 : 1 million scale for the whole of Europe, with progress towards harmonization. It utilized the geological datasets held by the various geological survey organizations in Europe. These datasets differ considerably with respect to their contents, descriptions, and geometries. To make those data interoperable was a considerable task, and 1G-E provided the basis for that endeavor. The 1G-E Geology Data Specification comprised core vocabularies based on the IUGS-CGI vocabularies [25] to describe the lithologies, ages, and genesis of rocks and tectonic structures. The vocabularies defined terms and the relationships between those terms.

Based on the IUGS-CGI GeoSciML, OneGeology-Europe employed a distributed digital model in which the data remained with the responsible geological survey organization. Each survey provided access to its data via the Internet, and the OneGeology-Europe portal (Fig. 27.20) harvested that data. Thus, each geological survey implemented and hosted an interoperable web service, delivering their national geological data in a semantically harmonized form.

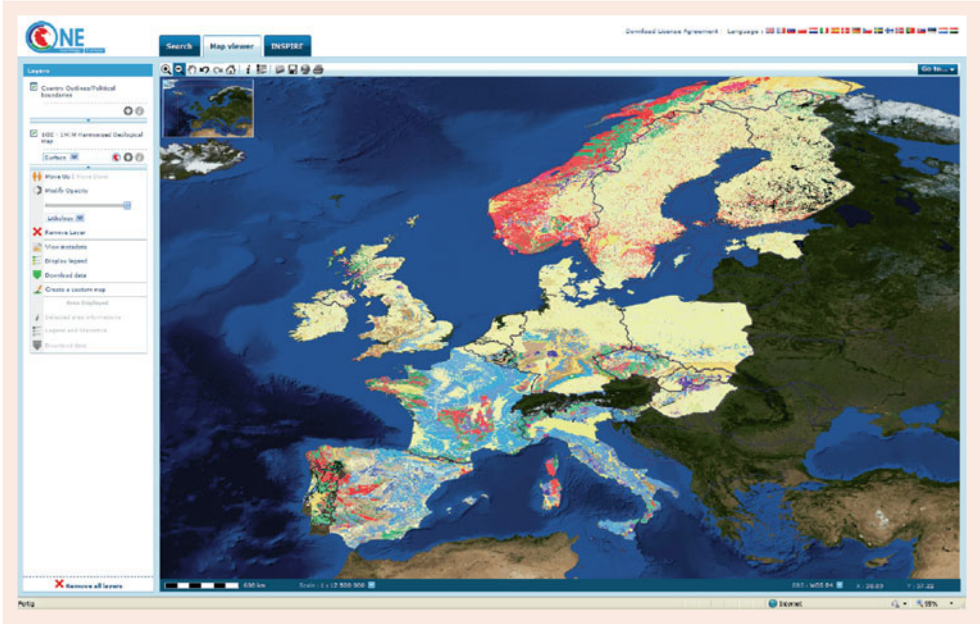


Fig. 27.20 The former portal of the EC project OneGeology-Europe. Distributed data from 20 geological survey organizations across Europe were visualized using and downloaded from this now-defunct website

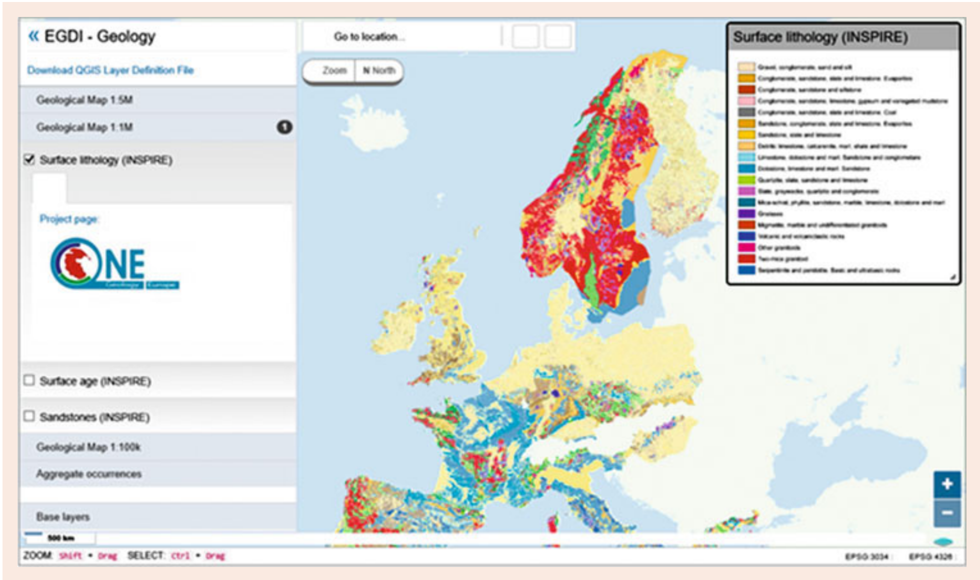


Fig. 27.21 OneGeology-Europe data are now served by the EGDI portal (http://www.europe-geology.eu/onshore-geology/geological-map/onegeologyeurope/)

These vocabularies were the basis for how the geological surveys participating in OneGeology-Europe described the geology of their country within the project and how they delivered that data in accordance with the agreed 1G-E data model and data specification. The original national databases remained unchanged.

Once the national datasets were provided to the project in accordance with this data specification, the team and surveys reviewed, noted issues with, and where possible *reworked* the datasets to progress towards geometric harmonization—a crucial step towards INSPIRE goals. The vocabularies and framework developed in that step could then be upscaled to more detailed levels and progressively deployed for higher-resolution geological data in a later step (Fig. 27.20; [22]).

The vocabularies, data model, and portrayal rules developed within OneGeology-Europe served as a prototype for the implementation of the INSPIRE Directive for the geology theme [22].

Today, 1G-E data are provided by the EuroGeoSurveys EGDI portal (see Fig. 27.21; [72]).

27.8.4 European Marine Observation and Data Network (EMODnet)

The EC EMODnet Geology project is one of seven projects that bring together relevant spatial information on the European marine environment: Geology, Chemistry, Biology,

Fig. 27.22 EMODnet Geology Portal: map of the lithology of the European seafloor [73]. The data used in this figure were made available by the EMODnet Geology project (see https://www.emodnet-geology.eu) funded by the European Commission Directorate for Maritime Affairs and Fisheries. These data were compiled by Kristine Asch, BGR, from the EMODnet Phase III partners

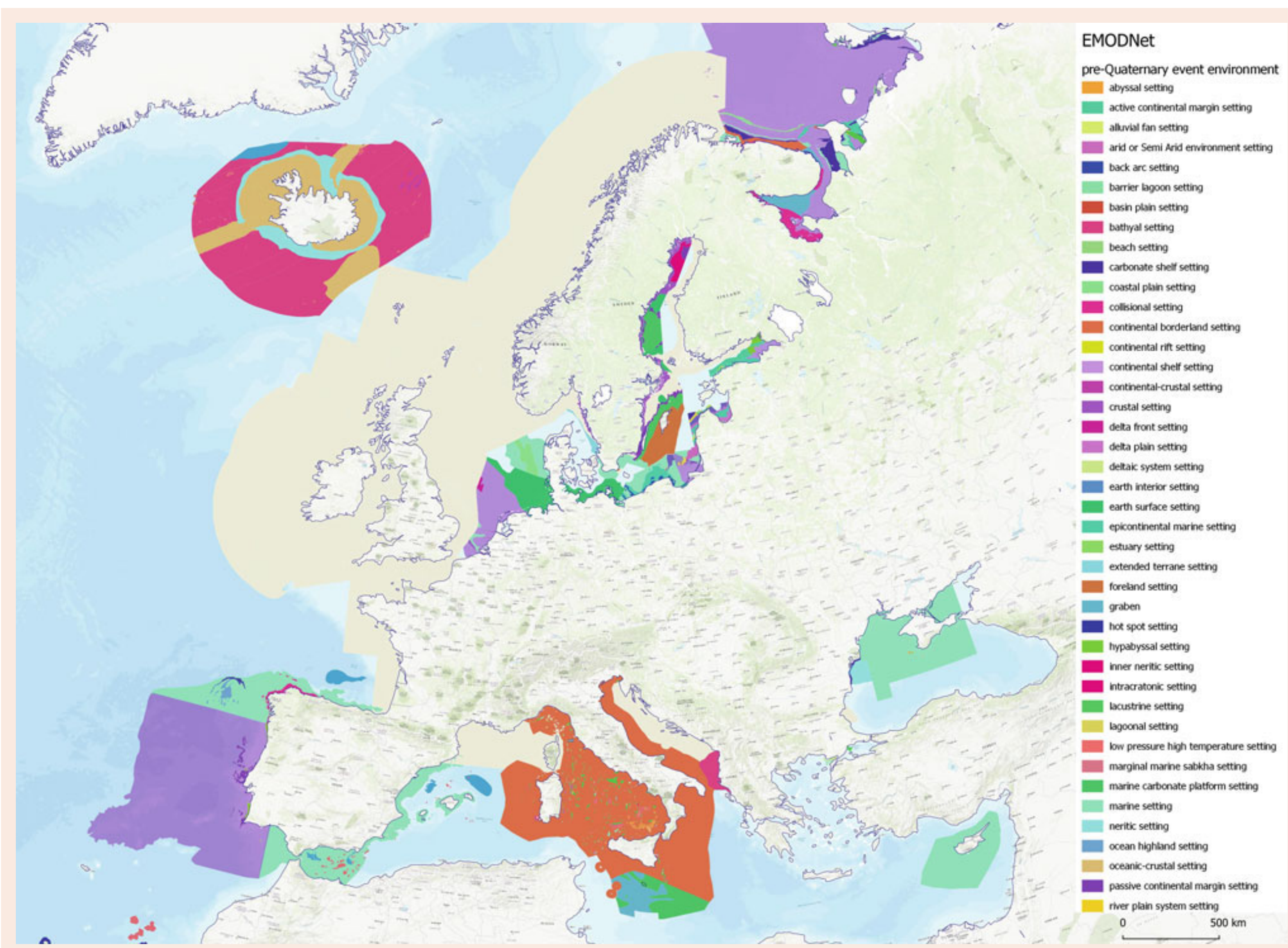


Fig. 27.23 EMODnet Geology Portal: map of rock genesis on the European seafloor: (event environment) [73]. The data used in this figure were made available by the EMODnet Geology project (see https://www.emodnet-geology.eu), funded by the European Commission Directorate for Maritime Affairs and Fisheries. These data were compiled by Kristine Asch, BGR, from the EMODnet Phase III partners

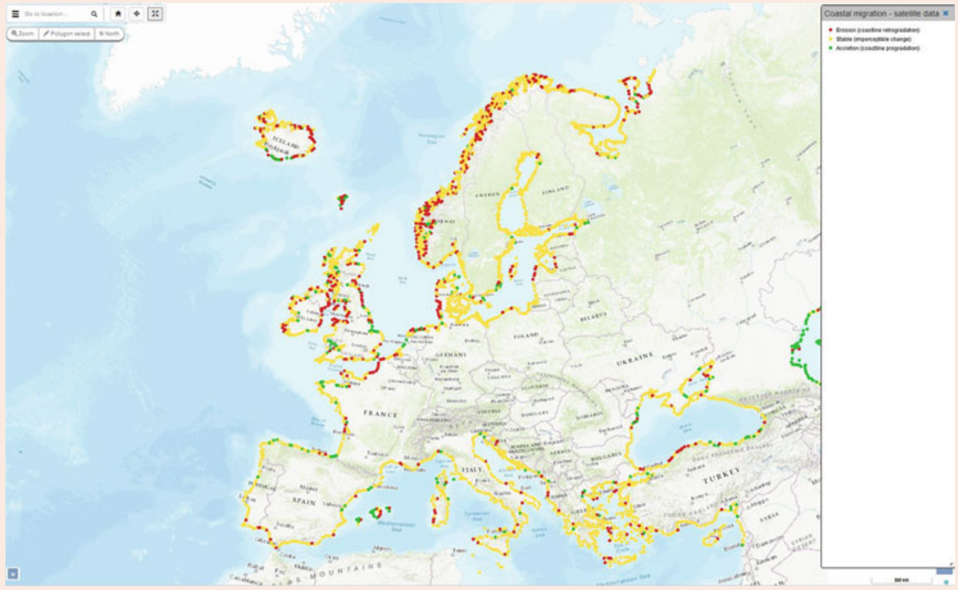


Fig. 27.24 EMODnet Geology Portal: map of coastal migration in Europe (status as of 8/2018) [73]. The data used in this figure were made available by the EMODnet Geology project (see https://www.emodnet-geology.eu), funded by the European Commission Directorate for Maritime Affairs and fisheries. These data were compiled by Sytze van der Heteren, TNO, from the EMODnet III partners

Physics, Bathymetry, Seabed Habitats, and Human Activities. Similar to OneGeology, the aim of EMODnet is to make hidden data interoperable, harmonize them as far as possible, and make those data publicly available. So far, there have been four project phases of EMODnet Geology.

During the first (preparatory) phase (2009–2012), 14 organizations from 14 European countries compiled and semantically harmonized geological information at the 1 : 1 million scale to demonstrate the feasibility of such an endeavor. The focus was on the identification of the geological information that existed in each country represented in the project and the construction of a new EMODnet Geology portal (Figs. 27.22, 27.23, 27.24).

In 2013, the second phase started, with 36 organizations (mainly geological survey organizations) from 30 countries compiling information on:

- Seabed substrates (sediment layers on the sea floor) and sediment accumulation rates
- Sea-floor geology: lithology (bedrock geology beneath the surficial sediment and Quaternary deposits), stratigraphy, and genesis (event environment and event process)
- Coastal behavior
- Mineral occurrences (e. g., oil and gas, aggregates, metallic minerals)
- Geological events and probabilities (e. g., earthquakes, submarine landslides, volcanic centers).

Spatial data were collated for European seas, including those off Iceland, Norway, and Russia as well as the Iberian Peninsula, Italy, Greece, and Turkey.

In spring 2017, the third phase of EMODnet began, in which the data on European seas were consolidated and enriched with more detail and content. Two new themes—geomorphology and submerged landscapes—were added, and specific data specifications for various themes—seabed substrates, sea-floor geology, coastal behavior, mineral occurrences, and geological events and probabilities (e. g., earthquakes, submarine landslides, and volcanic centers)—were developed and used by all participants to structure and transform their data.

The fourth phase started in September 2019 with additional partners, and aims to further dataset completion and consolidation within 2 years.

In contrast to OneGeology and OneGeology-Europe, geological data for each of the different themes within EMODnet are being compiled and harmonized within the topics seabed substrate, seafloor geology, coastal behaviour, events and probabilities, marine minerals and submerged landscapes. Only when one phase of compilation and harmonization is completed as best as possible are the data provided to the central portal, where they can be downloaded.

27.8.5 Geoscience Information in Africa

Most African countries have common geoscience information challenges, such as human resources (finding and retaining skilled staff), isolated projects, and difficulties in accessing data and research results and managing geoscience data and infrastructure.

The Geoscience InfoRmation in AFrica network (GIRAF) was founded in 2009 [74] during a conference workshop in Namibia under the umbrella of the IUGS CGI and UNESCO to provide a forum to address and improve the above situation. Meanwhile, through a series of three more workshops and interspersed smaller events, GIRAF has built a platform for African geoscience information experts to share information and experience and to cooperate and network across political boundaries. GIRAF also facilitates the identification of knowledge gaps and brings African perspectives to the global geoscience community.

At the GIRAF 2009 Workshop, delegates from 26 African nations and 4 European countries as well as representatives from UNESCO, ICSU, and IUGS-CGI met and pursued five aims:

1. To bring together relevant African authorities, national experts, and stakeholders in geoscience information
2. To initiate the creation of a pan-African geoscience information knowledge network to exchange and share geoscience information, knowledge, and best practices
3. To integrate authorities, national experts, and other experts across Africa into global geoinformation initiatives
4. To develop a strategic plan for Africa's future in geoscience information
5. To make Africa a more active part of the international geoscience information community.

The program for the GIRAF 2009 workshop was designed to explore each of these aspects aimed at enhancing the application of geoscience information to improve the health and prosperity of the people in Africa.

GIRAF concentrates on fostering the provision, dissemination, and use of geoscience information and geodata in Africa through sound geodata management supporting sustainable development in the fields of mineral planning and mining, artisanal mining (environmental impact studies, legalization, health aspects, socially acceptable living conditions in mining housing estates, etc.), securing and protecting water resources, and soil protection.

GIRAF is bringing together scientists from over 30 African countries at three other conference workshops in Tanzania, Ghana, and Mozambique and at several events between those workshops, and has grown into a body comprising around 400 members. In 2016, a constitution for GIRAF was agreed upon, and an office for GIRAF was set up in the African Minerals and Geosciences Centre in Dar es Salaam (Tanzania). The GIRAF network was finally transferred into the hands of African colleagues at the 26th Colloquium of African Geology (CAG26) in Ibadan, Nigeria.

27.8.6 Sensor-Based Landslide Early Warning System (SLEWS)

SLEWS (A Sensor-based Landslide Early Warning System) is a joint project aimed at the systemic development of a prototyping alarm and early warning system for different types of landslides utilizing ad-hoc wireless sensor networks and spatial data infrastructure technologies according to OGC guidelines for real-time monitoring.

SLEWS consists of different kinds of subsystems, such as monitoring, information, warning, and alarm systems. The monitoring system is based on a wireless sensor network (WSN) and enables environmental data to be requested and accessed in near real time within an information system. With the warning and alarm system, notifications are forwarded to decision makers when thresholds are exceeded [75].

Data Capture and Information Distribution

Data capture is organized by an ad-hoc and autonomous wireless sensor network (see Chap. 9). Information retrieval results from data and sensor fusion. The information is provided by standardized interfaces complying with the OGC Sensor Web Enablement (SWE) specifications for user-orientated processing and visualization by a spatial data infrastructure (SDI). The open structure of the system allows very rapid and flexible adjustment to changes in boundary conditions and permits simple linkage to other data sources or sensor networks.

Due to the progressive development of urban areas and infrastructure, more and more people are settling in environments that become or already are endangered by mass movement. This situation is complicated by the dependency of today's society on functioning infrastructure and the increasing number of human and/or built objects in endangered areas. SLEWS is based on an ad-hoc WSN for landslide monitoring and on the application of innovative service-based web technologies integrating OGC standards for web services and retrieval. Conventional monitoring systems for early warning are cost-intensive monolithic systems.

The application of a WSN and highly flexible sensors allows for a very flexible, easy to set up, secure, and cost-efficient monitoring system. The intelligent integration of different sensors may be used to improve data quality and function control by sensor fusion. At a larger scale, intelligent sensor configuration and positioning can be used to cross-validate information by network fusion. Data from foreign providers can be integrated due to an open platform strategy using OGC Web Processing Services (WPS).

The prototype service based on a Spatial Data Infrastructure (SDI) involved sensors, geodata, recent information and communication, and methods and models to estimate parameters with relevance to certain areas, e. g., to landslides. The new approach allows any involved institution to obtain user-adapted information at a very early stage independently of the hierarchical information structure. The improvement of user interfaces and the possibility of model and prognosis integration are of major interest and will lead to the development of new standards [51, 76].

OGC Web Services

Using the web services of Sensor Web Enablement, scientific authorities can request information from different sensors via the Internet. The Sensor Observation Service (SOS) also provides bidirectional communication. Therefore, data can be transferred from the system to the operators. Actual measurements and observational data can be exchanged by the XML-based OpenGIS Observations and Measurements Encoding Standard (O&M). In the case of slow-moving landslide processes, graduated warning messages can be sent automatically to different recipients via the Web Notification Service (WNS). Users must subscribe in advance for the Sensor Alert Service (SAS). Within the alarm system, messages are forwarded via the WNS. Thus, transportation lines (railway lines or highways) can be closed in time, e. g., by SMS-based traffic light operation.

27.9 Future Challenges

It is very apparent that around the world, in universities and in geological surveys and other agencies, there are very similar goals and activities, issues, and challenges in geoscience information. The rapid development of technology and especially web services, the massive explosion of data, and the increasingly diverse requirements of users across all sectors—governments, scientists, commerce, and the public—all introduce new demands and challenges. When we add in global and transnational issues where geoscience has a critical role to play, such as sustainable energy resources, mineral resources, agriculture, groundwater, transport, natural hazards such as earthquakes and tsunamis, and last but not least climate change, it becomes obvious that there is a pressing need for cooperation to sustain future generations on Earth. While geoscience informatics has many good examples of the exploitation of developments in information technology to further global cooperation, we still have a long way to go in sharing information, experience, and expertise efficiently so that we can analyze and synthesize data and knowledge across political and continental boundaries with minimal barriers, work online in international teams on research projects, and add value for future generations.

The Earth sciences are currently in a state of radical change. The last decade has seen a paradigm shift from studying our planet as a sum of discrete components to adopting an integrated systems approach. The concept of Earth system science has emerged as a more natural, unified way of studying past and present processes on Earth. Dramatic increases in human population, resource consumption, and environmental degradation coupled with socioeconomic concerns and the concept of sustainable development have redefined the role of the Earth sciences in our society. In particular, there is a need to assess and predict the consequences of anthropogenic forces for natural Earth systems.

In recent centuries, an immense treasure trove of (often fragmented) geoscience data has been accumulated in geological organizations, and so an essential task for the future is to process, harmonize, analyze, make use of, and preserve those data for future generations. This necessitates communication and international and transorganizational cooperation, which has already led to a stable set of commonly agreed standards, directives, and other types of agreements. The previous sections have illustrated the multitude of worldwide technical activities in this context, of which the most prominent international examples are INSPIRE, OneGeology, and EMODnet. Consequently, the near-future challenges may be characterized by the following statements;

- Communication between groups, organizations, and projects working in the area of SDI and geoscience information needs greater cohesion
- Higher priority and more resources should be devoted to cooperative, synergetic goals, e. g., transboundary research data infrastructures
- Further development of international standards, including input from developing countries, is needed
- The challenges of Big Data need to be faced
- Further development of 3-D/4-D modeling must be pursued
- Considerable effort needs to be put into processing, harmonizing, analyzing, and preserving existing data
- Realizing potential synergies and reducing overlaps and duplication will not happen if left to individual groups
- Domain-specific associations such as EuroGeoSurveys, EuroGeographics, and IUGS should play a stronger role
- There is a need for increased proactive orchestration of the geoscience informatics and broader SDI communities, and for the provision of funding streams that encourage integration, not fragmentation
- More effort and resources in SDI projects should be devoted to communication and outreach.

27

References

1. Charpentier, J.W.F.: The Map of the entire Transbaikal region produced by and the Geological Map of England Wales from William "Strata" Smith establishing the principles of chronostratigraphy and biostratigraphy (1778)
2. Butcher, N.E.: The advent of colour-printed geological maps in Britain. In: Proc. R. Inst. G. B., vol. 55, pp. 149–161. (1984)
3. Winchester, S.: The Map That Changed the World – The Tale of William Smith and the Birth of a Science. Viking, London (2001)
4. Zitzmann, A., Lepper, J.: Geologische Übersichtskarte 1:200000 (GÜK 200). BGR, Hannover (1979). Blatt CC4718
5. Loudon, T.V.: Geoscience After IT. A View of the Present and Future Impact of Information Technology in Geoscience. Computer Methods in Geosciences, vol. 17. Pergamon, Kidlington (2000)

6. Maltman, A.: Geological Maps – An Introduction, 2nd edn. Wiley, Chichester (1998)
7. Asch, K.: GIS-gestützte Standortfindung für Siedlungsabfalldeponien. In: Asch, K. (ed.) GIS in Geowissenschaften und Umwelt, pp. 63–71. Springer, Berlin, Heidelberg (1999). in German
8. Bonham-Carter, G.F.: Geographic Information Systems for Geoscientists, Modelling with GIS. Computer Methods in the Geosciences, vol. 13. Pergamon, Kidlington (1994)
9. Noble, R.R.P., Gonzalez-Alvarez, I., Reid, N., Krapf, C.B.E., Pinchand, T., Cole, D., Lau, I.: Regional Geochemistry of the Coompana Area. CSIRO, Perth (2018). Tech. Rep. EP187470
10. Bickmore, D.P., Kelk, B.: Production of a multi-coloured geological map by automated means. In: Proc. 24th IGC, Ottawa, pp. 121–127. (1972)
11. Asch, K., Bavec, M., Bergman, S., Carter, M., Coupek, P., Colbach, R., Declercq, P.-Y., González Clavijo, E., Gruijters, S., Gürtlerova, P., Hennings, S., Janjou, D., Kacer, S., Klicker, M., Nironen, M., Novak, M., Orosz, L., Pantaloni, M., Pielach, M., Ploom, K., Cerdan, F.P., Riisager, P., Schubert, C., Schokker, J., Sikhegyi, F., Smith, A., Stepién, U.: OneGeology-Europe. Scientific, Semantic and Geometric Harmonisation of Spatial Geological Data in Europe – Issues and Recommendations (2010). ECP-2007-GEO-317001
12. Asch, K.: The 1:5 Million International Geological Map of Europe and Adjacent Areas: Development and Implementation of a GIS-Enables Concept. BGR, Hannover (2003)
13. ISO 710-1:1974: Graphical Symbols for Use on Detailed Maps, Plans and Geological Cross-Sections – Part 1: General Rules of Representation
14. ISO 710-2:1974: Graphical symbols for use on detailed maps, plans and geological cross-sections – Part 2: Representation of Sedimentary Rocks
15. ISO 710-3:1974: Graphical Symbols for Use on Detailed Maps, Plans and Geological Cross-Sections – Part 3: Representation of Magmatic Rocks
16. ISO 710-4:1982: Graphical Symbols for Use on Detailed Maps, Plans and Geological Cross-Sections – Part 4: Representation of Metamorphic Rocks
17. ISO 710-5:1989: Graphical Symbols for Use on Detailed Maps, Plans and Geological Cross-Sections – Part 5: Representation of Minerals
18. ISO 710-6:1984: Graphical Symbols for Use on Detailed Maps, Plans and Geological Cross-Sections – Part 6: Representation of Contact Rocks and Rocks Which Have Undergone Metasomatic, Pneumatolytic or Hydrothermal Transformation or Transformation by Weathering
19. ISO 710-7:1984: Graphical Symbols for Use on Detailed Maps, Plans and Geological Cross-Sections – Part 7: Tectonic Symbols
20. ISO 14688-2:2017: Geotechnical Investigation and Testing – Identification and Classification of Soil – Part 2: Principles for a Classification
21. ISO 14689-1:2003: Geotechnical Investigation and Testing – Identification and Classification of Rock – Part 1: Identification and Description
22. INSPIRE Thematic Working Group Geology: D2.8.II.4 Data Specification on Geology – Technical Guidelines. EC-JRC, Ispra (2013)
23. Cohen, K.M., Harper, D.A.T., Gibbard, P.L.: ICS International Chronostratigraphic Chart 2018/07 (2018). www.stratigraphy.org, Accessed 23 July 2018
24. OGC: Geoscience Markup Language (GeoSciML). https://www.opengeospatial.org/standards/geosciml, Accessed 15 Feb 2020
25. International Union of Geological Sciences – Commission on the Management and Application of Geoscience Information (IUGS-CGI). http://www.cgi-iugs.org/, Accessed 1 May 2022
26. ISO 19136-1:2020: Geographic Information – Geography Markup Language (GML) – Part 1: Fundamentals
27. ISO 19142:2010: Geographic Information – Web Feature Service
28. ISO 19156:2011: Geographic Information – Observations and Measurements
29. North American Geologic Map Data Model Steering Committee: NADM Conceptual Model 1.0 – A conceptual model for geologic map information: US Geological Survey Open-File Report 2004-1334 (2004). available from http://pubs.usgs.gov/of/2004/1334, accessed 9 Jan 2019; also published as Geological Survey of Canada Open File 4737, 1 CD-ROM (2004)
30. U.S. Geoscience Information Network (USGIN). http://www.usgin.org/, Accessed 15 Feb 2020
31. AuScope. https://www.auscope.org.au/, Accessed 15 Feb 2020
32. Australian Geoscience Information Network (AUSGIN). http://www.geoscience.gov.au/, Accessed 15 Feb 2020
33. EarthResourceML (ERML). http://www.earthresourceml.org/, Accessed 15 Feb 2020
34. International Union of Geological Sciences (IUGS). http://www.iugs.org/, Accessed 15 Feb 2020
35. Sensor Web Enablement (SWE) Common Data Model Encoding Standard. https://www.opengeospatial.org/standards/swecommon, Accessed 15 Feb 2020
36. EURARE (2019). http://www.eurare.eu/, Accessed 9 Jan 2019
37. Minerals4EU (2019). http://www.minerals4eu.eu/, Accessed 9 Jan 2019
38. GeoScienceML (2019). http://geosciml.org, Accessed 1 May 2022
39. Anderson, C.L.: Topographic Representation or Model. Patent US298812 A (1884). issued 20 May 1884, http://www.google.us/patents/US298812
40. Barringer, D.M.: Apparatus for Illustrating Geological Formations. Patent US477633 A (1892). filed 15 Jun 1891, and issued 28 Jun 1892, http://www.google.us/patents/US477633
41. Loop Consortium. https://loop3d.org/, Accessed 15 Feb 2020
42. Jackson, I.: Addressing the real needs of all the users of geological information: The opportunities, issues and problems. In: The Current Role of Geological Mapping in Geosciences NATO Sci. Ser. IV: Earth and Environmental Sciences, vol. 56, pp. 59–68. (2005)
43. Wellmann, J.F., Regenauer-Lieb, K.: Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models. Tectonophysics **526–529**, 207–216 (2012)
44. Perrin, M., Zhu, B., Rainaud, J.F., Schneider, S.: Knowledge-driven applications for geological modelling. J. Petrol. Sci. Eng. **47**, 89–104 (2005)
45. Rosenbaum, M.S., Turner, A.K. (eds.): New Paradigms in Subsurface Prediction. Lecture Notes in Earth Sciences, vol. 99. Springer, Berlin (2003)
46. Lindsay, M.D., Jessell, M.W., Ailleres, L., Perrouty, S., de Kemp, E., Betts, P.G.: Geodiversity: Exploration of 3D geological model space. Tectonophysics **594**, 27–37 (2013)
47. Jessell, M.W., Ailleres, L., de Kemp, E.A.: Towards an integrated inversion of geoscientific data: What price of geology? Tectonophysics **490**(3), 294–306 (2010)
48. Bond, C.E., Shipton, Z.K., Jones, R.R., Butler, R.W.H., Gibbs, A.D.: Knowledge transfer in a digital world: Field data acquisition, uncertainty, visualization, and data management. Geosphere **3**(6), 568–576 (2007)
49. Turner, A.K., Gable, C.W.: A review of geological modelling. In: Three-Dimensional Geologic Mapping for Groundwater Applications, pp. 75–79. (2007). Open-file Report 07-4, Minnesota Geological Survey
50. GSI3D (2019). http://www.gsi3d.org.uk, Accessed 9 Jan 2019
51. Berg, R.C., Mathers, S.J., Kessler, H., Keefer, D.: Synopsis of current 3D geological mapping and modeling in geological survey organizations. Ill. State Geol. Surv. Circ. **578**, 92 (2011)

52. BGR: SLEWS – A Sensorbased Landslide Early Warning System (2019). https://www.bgr.bund.de/EN/Themen/Sammlungen-Grundlagen/GG_geol_Info/Projekte/abgeschlossen/SLEWS/slews_inhalt_en.html, Accessed 9 Jan 2019
53. Mallet, J.-L.: GeoModelling. Oxford Univ. Press, New York (2002)
54. Zanchi, A., DeDonatis, M., Gibbs, A., Mallet, J.-L.: Imaging geology in 3D. Comput. Geosci. **35**, 1–3 (2009)
55. GOCAD (2019). https://www.pdgm.com/products/GOCAD, Accessed 1 May 2022
56. GeoModeller (2019). http://www.geomodeller.com, Accessed 1 May 2022
57. Cowan, E.J., Beatson, R., Ross, H.J., Fright, W.R., McLennan, T.J., Mitchell, T.J.: Rapid geological modelling. In: Vearncombe, S. (ed.) Appl. Struct. Geol. Miner. Explor. Min. Int. Symp., Kalgoorlie Australian Institute of Geoscientists Bulletin, vol. 36, pp. 89–99. (2002). https://www.aig.org.au/publication-shop/digital-aig-bulletin-no-36-applied-structural-geology-for-mineral-exploration-and-mining/
58. Terrington, R.L., Mathers, S.J., Kessler, H., Hulland, V., Price, S.J.: [wk3] Subsurface viewer 2009. User manual V1.0. (OR/09/027) (2009). http://nora.nerc.ac.uk/7195/1/Subsurface_Viewer_2009_UserManual_v1.pdf, Accessed 9 Jan 2019. British Geological Survey Internal Report
59. British Geological Survey (2019). https://www.bgs.ac.uk/datasets/uk3d/, Accessed 1 May 2022
60. Stern, N.: Stern Review on the Economics of Climate Change. HM Treasury, London (2006). available from http://www.hm-treasury.gov.uk/sternreview_index.htm, accessed 9 Jan 2019
61. Pachauri, R.K., Reisinger, A. (eds.): Climate Change 2007. IPCC, Geneva (2007). Synthesis Report, Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change Core Writing Team
62. Hughes, A., Graham, M., Jackson, C., Mansour, M., Vounaki, T.: ZOOM into GSI3D. In: Ext. Abstr. 2nd Int. GSI3D Conf. pp. 28–29. BGS, Keyworth (2008)
63. Hadlow, N., Molyneux, I., Gallagher, A., Robelin, C.: The development of chalk catchment ground models in Southern England and Northern France. In: Ext. Abstr. 2nd Int. GSI3D Conf. pp. 24–25. BGS, Keyworth (2008)
64. Kessler, H., Mathers, S.J., Lelliott, M., Hughes, A., MacDonald, D.: Rigorous 3D geological models as the basis for groundwater modelling. In: Three-Dimens. Geol. Mapp. Groundw. Appl. Workshop Ext. Abstr., Denver (2007). available from http://nora.nerc.ac.uk/4129/1/kessler.pdf, accessed 9 Jan 2019
65. British Geological Survey: A vision for the BGS Environmental Modelling Platform. https://www.bgs.ac.uk/about/vision.html, Accessed 15 Feb 2020
66. CSIRO: Coastal Environmental Modelling. https://research.csiro.au/cem/csiro, Accessed 15 Feb 2020
67. Asch, K.: New EC directive to "INSPIRE" the European Geoscience Community. Environ. Geol. **57**, 959–961 (2009)
68. INSPIRE: Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)
69. INSPIRE: Roadmap (2019). https://inspire.ec.europa.eu/inspire-roadmap/, Accessed 9 Jan 2019
70. OneGeology – Portal website: http://portal.onegeology.org/ (Languages: English and French); General website: www.onegeology.org (Languages: English, introductory pages in French, Chinese, Russian, Spanish and Arabic); Website: http://www.bgr.bund.de/OneGeology. Accessed 9 Jan 2019
71. OneGeology (2019). https://onegeology.org, Accessed 1 May 2022
72. EGDI (2019). http://www.europe-geology.eu/onshore-geology/geological-map/onegeologyeurope/, Accessed 9 Jan 2019
73. EMODnet Geology Portal (2019). http://www.emodnet-geology.eu/map-viewer/, Accessed 9 Jan 2019
74. British Geological Survey: GIRAF (2009). http://giraf-network.net/index.php/en/workshops/gi-raf-2009-work-shop.htm, Accessed 1 May 2022
75. Fernández-Steeger, T.M., Walter, K., Haß, S., Niemeyer, F., Nakaten, B., Homfeld, S.D., Asch, K., Azzam, R., Bill, R., Ritter, H.: SLEWS – A prototype system for flexible real time monitoring of landslides using an open spatial data infrastructure and wireless sensor networks. In: Early Warn. Syst. Earth Manag. Status Semin. pp. 3–15. (2009). 12–13 Oct 2009, TU München; Programme & Abstracts
76. Haß, S., Arnhardt, C., Asch, K., Fernandez-Steeger, T.: SLEWS: A people-centred landslide early warning system in the context of risk management. In: Forum Katastrophenvorsorge: Katastrophen – Datenhintergrund und Informationen, Bonn (2009). Poster 10, in German

Kristine Asch Kristine Asch, PhD, is a geologist and leads the unit "Geological Information Systems and Maps" at the German Federal Institute for Geosciences and Natural Resources. She is Vice-president of the European Subcommission of the Commission of the Geological Map of the World. Kristine contributes to national and international geological standards and classifications such as the INSPIRE data specifications or the IUGS geoscience terminology. She coordinates and builds international GIS-Projects such as the International Geological and Quaternary Maps of Europe (IGME 5000, IQUAME 5000) and the EMODnet Geology project series.

Jens Klump Jens Klump received his PhD in Marine Geology from the University of Bremen in 1999. He worked at the Helmholtz Center for Geoscience Potsdam GFZ before moving to the Mineral Resources unit of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Perth, Western Australia. His work focuses on the application of information technology to geological research.

Stephen J. Mathers Steve retired from The British Geological Survey in 2016 after 38 years with the organization. He formerly headed up a team involved in the construction of a National Geological Model for the United Kingdom and the development of geological modeling at BGS.

Holger Kessler Holger Kessler joined the BGS with a degree in Physical Geography from Frankfurt University as a Survey Geologist in 1998. After a career in primary geological mapping and modeling in Yorkshire and Norfolk, he was instrumental in shaping the BGS' modeling strategy until 2018; he is now seconded to central government developing the UK's geospatial strategy.

Geographic Information Systems in Energy and Utilities

28

William Meehan, Robert G. Brook, and Jessica Wyland

Contents

W. Meehan (✉)
Utility Solutions, Esri
Redlands, CA, USA
e-mail: bmeehan@esri.com

R. G. Brook (✉)
Hecatic Labs
Los Angeles, CA, USA
e-mail: rob@hecaticlabs.com

J. Wyland
Marketing Communication, Esri
Redlands, CA, USA
e-mail: jwyland@esri.com

Abstract

Utilities and energy companies have used GIS for decades. However, their primary use has been as a system of record, mostly for electric and gas distribution and to a lesser extent for transmission assets. In the energy production business, GIS has been used primarily for exploration activities. Up until recently, GIS professionals have been the main users. Most utility workers see GIS as powerful but not generally suitable for casual users. The technological barriers have been removed. Even though GIS has powerful geospatial and analytical capabilities, it embraces commercial technology made popular by social media and mobile devices. Further, energy companies are beginning to incorporate real-time data into their GIS. Finally, while people associate GIS with two-dimensional maps, companies are beginning to view the spatial information in their GIS in 3-D.

The emergence of the GIS platform has liberated the rich data and functionality that heretofore was only available to the GIS professional. One of the common complaints from energy companies is the lack of currency of their data. By adopting a location platform based on GIS technology, energy companies can now access, update, report, and visualize the data from their mobile devices, thus shrinking the data latency from days, weeks, and even months (sometimes even years) to seconds.

Traditional GIS is still a critical information technology for energy companies and utilities. The GIS platform (as opposed to the legacy desktop-only GIS) multiplies that value. Energy companies deal with assets, people, and processes with strong spatial dimensions. Utilities have assets and customers distributed nearly everywhere in the world. Gas companies have pipes up and down city streets. Oil rigs are spread throughout the world. The work of energy companies and utilities is always linked to geography. That is why the GIS platform makes such a powerful business solution.

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_28

This chapter will focus on how the role of the GIS platform has evolved over the last several years. It will show how GIS help energy companies better manage their work, assets, and people. It will illustrate how GIS has evolved from a niche application used by only a few people to an enterprise system used by nearly everyone in the organization. The data managed by the GIS is widely available.

Keywords

System of record · System of insight · System of engagement · Universal access · GIS · SCADA · ADMS · Tracing · Utility · Pipeline · Network Model

28.1 GIS is now a Platform

For decades, GIS has been used extensively by professionals in the oil, gas, and utilities industry. While the powerful analytic capabilities of the platform are used to some extent, the vast majority of uses of GIS have been as what industry people call a *system of record*. That is, GIS is used as a very convenient means of locating an organization's assets. Historically, the utilities industry often called GIS an *automated mapping/-facilities management system* or AM/FM. Even today, people in industry call GIS by its old name. For many, GIS was a very good technology for automating the old manual maps they had been drawing for years. This focus on the replication of hand drawn maps led many to limit the broad and transformative aspects of the capabilities of today's modern GIS.

28.1.1 What is the Platform?

That is changing. This shift is due to the acceptance of and common use of technology platforms. In his seminal book *Age of the Platform*, author *Phil Simon* implies what a platform does. It allows people to reach and connect with one another and obtain information. The operative word here is *connected*. Modern social media, entertainment, and retail platforms such as Facebook, Instagram, Snapchat, Apple iTunes, Amazon, and Google do this routinely.

Rather than focusing on creating a map that looks very much like the old operating sheets or transmission strip plans, with text placement and white-space management being a prominent component of the technology, a modern GIS platform provides the ability to connect with people (workers, contractors, the media, first responders, customers, and regulators) in much the same way that Facebook communicates with its billion subscribers. Further, the platform does not really differentiate according to the user's device. The main objective of a platform, as Simon proclaims, is the ability to connect people to get information. For many energy companies, that communication medium is still printed maps.

The modern GIS breaks that mold. Employees still use a map to find where their assets are. Next, they need to validate that the information about the asset is correct. Finally, they have to correct any errors they find quickly. They must get this information in the hands of a wide variety of workers. This old way contrasts this. Mark corrections on a paper copy or printout of the map. Send it to the office to update their records. Then sometime later, print out a new map. This all flies in the face of the modern concept of a platform. Platforms do this in near real-time. The notion of immediate updates of information from everywhere provides immediate access to all critical information for everyone involved. This means that information is available from the company's boardroom to the dashboard on any device anywhere at any time.

For utilities, there are actually three broad capabilities that the GIS platform can provide. These capabilities all start with the letter c. They are:

- Communication—the ability to communicate using location data with virtually anyone within the company and, if authorized, any person outside the company, such as first responders, the media, or government officials.
- Collaboration—the ability to share asset and location information interactively with other workers, contractors, first responders, and others in the office and the field
- Coordination—the ability to show a variety of tasks from different entities (such as electric, gas, pipelines, and telecommunication facilities) in the same display and perform a spatial analysis to determine interferences or safety issues.

There are actually many complex work practices that require many participants to work together. Often those participants are in different locations. Knowing what each is doing is critical to avoiding costly mistakes and safety issues. For example, when a city plans to repave a street but does not coordinate that work in real time with utilities, the possibility exists that the utilities will dig up newly paved streets. Communication, collaboration, and coordination require that all interested parties work in harmony. Location plays a huge role.

28.1.2 Elements of a GIS Platform

Today's platforms have some common elements. These elements differentiate an application from a platform. An application abstracts a workflow. A platform is an information infrastructure. A GIS platform provides more than the functionality of a GIS application. GIS applications, however, operate on the GIS platform infrastructure, rendering them more powerful than ever. The following are the elements of a GIS platform:

1. Identity: Just like familiar consumer platforms, a platform knows who is using it. A platform understands that a particular user is the same user even when they move devices.
2. Device independence: As noted above, platforms allow users to move from one kind of device to another freely. While the user experience will vary from one device to another, the core functionality remains constant.
3. Leverages external data: This is one of the most valuable aspects of a platform. For example, a GIS platform can leverage data from the web about demographics, the real-time weather, traffic patterns, buying behavior, and construction activity. The list is almost endless. This provides a superb platform for smart communities. What makes a GIS platform different from other platforms is the focus on location.
4. The ability to cooperate with other platforms: For example, a GIS platform cooperates with billing systems, real-time control systems, social media platforms, and virtually any platform that adopts common platform technology, such as web services.
5. Allows crowdsourcing: Users are allowed to add their observations and data to the platform. This adds a richness for energy company operations. For example, a citizen smells gas. That person then tweets about it. The gas company locates that tweet within their GIS platform, correlates that data with other similar tweets, and associates that collection of tweets with the company's infrastructure data in its GIS. The gas company then matches that location with known leaks and takes appropriate action.
6. The GIS platform is always on: Like most platforms, the GIS platform is designed to be always available.
7. Scaling: A platform must scale automatically as the demand on it increases. That is why platforms leverage cloud technology to a large degree. Likewise, a modern GIS platform does this, but it also provides the ability to preserve the privacy and security of its data. Hybrid cloud and on-premise systems are therefore common in the energy business.
8. Security built in: Platforms must be built to withstand cyber attacks, so security is built in.

28.2 Overall Picture

28.2.1 Serious Challenges for Energy Companies and Utilities

Energy companies and utilities face enormous challenges as they complete the first two decades of the twenty-first century. We have recognized society's complete and utter dependence on energy. Oil, gas, electricity, coal, and renewable energies drive economies worldwide. An interruption in any of these critical supply systems results in economic turmoil. It is also clear from both a scientific and political perspective that we cannot sustain the status quo in energy supply. Energy companies and utilities will need to change, and change fast.

While change introduces solutions to current problems, it often brings with it a brand-new set of problems to solve; for example, the proliferation of wind farms has created an energy supply free of greenhouse gas emissions at a reasonable cost, yet wind farms have been responsible for the deaths of thousands of birds, and some scientists warn of the eventual extinction of many bird species because of wind farms. Today, the power grid is secured by large, base-loaded generating units that stabilize the electricity supply. As wind generation increases, the stability of that supply becomes less certain due to the intermittent nature of wind. Hence, one solution leads to a new set of problems.

An oft-sought solution to the problems of energy companies and utilities, geographic information system (GIS) technology, can improve the production of an oil well, the distribution of electricity, and the transmission of natural gas. It cannot, however, directly stop wind farms from killing birds, or make solar energy more affordable. GIS will provide a spatial context for averting and/or solving the energy problems that are tied to location:

- Where do we drill a well?
- Which wells are producing?
- Where should we develop geothermal and wind resources?
- How can we run a transmission line from a proposed hydroelectric facility to the grid?
- Where do we place sensors on the electric transmission and distribution system to create a smart grid?
- Where do we run telecommunication systems to best communicate with smart meters?

The list goes on. GIS helps solve problems by making it simpler to discover relationships through spatial data; for example, if an oil-filled transformer fails, GIS can show you how long it will take for the spilled oil to reach a river, stream, or wetland. With a GIS, organizations can quickly assess concerns such as:

- What is the population density and demographic makeup of a particular area for a new gas station location?
- What is the relationship between customer satisfaction and recent gas company field crew locations?

GIS helps energy companies and utilities better understand the needs and drivers of their communities, customers, regulators, and financial stakeholders.

Now that the modern GIS embraces platform technology, users and interested stakeholders get immediate access to the rich intelligence the GIS provides. In fact, there are three a's that describe how this works. The GIS provides:

1. Access: Up until recently, GIS data, network functionality, results of spatial analysis, and data management have been largely limited to GIS professionals. While some data sharing by means of file transfers and pdfs are routine, data access to the core GIS capability has been limited. Platform technology changes all that. In effect, the platform liberates this intelligence and delivers it to anyone. Often, the GIS is referred to as a system of engagement, which means that the system of record is accessed throughout the enterprise.
2. Awareness: Now that the GIS platform can collaborate with other platforms, collect an endless supply of data from the web, stream real-time data, and participate in crowdsourcing, the GIS can much more fully answer the question: what's going on right now? For any company that struggles with emergency situations such as massive power failures, natural gas explosions, and large oil leaks, real-time situational awareness is essential to response and recovery efforts. This is also part of the system of engagement.
3. Analytics: This is at the heart of a true GIS. Layering the analytic capability onto platform infrastructure provides a timely understanding of the why and where of what might happen and allows energy companies to better plan and to provide timely decisions. The GIS is also referred to as a system of insight.

28.2.2 GIS Provides Spatial Context for Solutions

Energy companies and utilities operate in a world with many problems—population growth, climate change, social conflicts, resource shortages, and loss of biodiversity. Additionally, companies must cope with an aging infrastructure and workforce. The increasing complexity and severity of these problems portends a challenging future for society, particularly for energy organizations. These organizations have to figure out how to do more with less. That means fewer people with more work to do. This will happen through a voluntary exodus of people, or perhaps because of staff cuts. The organizations will need to be creative in finding new sources of revenue while reducing costs. Finally, they will need to adapt to ever-changing business, social, and global challenges, such as carbon trading and more stringent regulations. Clearly, we need better knowledge and awareness of our situation. We also need a more comprehensive approach to how we design and manage activities. We need an approach that considers and accounts for the impacts of our actions and guides us toward a more sustainable future.

Spatial knowledge provides a new way of thinking and problem solving by integrating geographic information into the systems we use to understand and manage assets, employees, and stakeholders. This approach allows us to *create* geographic knowledge by measuring, organizing data, and analyzing/modeling various processes and their relationships. This approach also allows us to *apply* this knowledge to the way we design, plan, and change our businesses.

28.2.3 Spatial Approaches Can Frame New Problems

GIS is an information system technology and now a platform with geography at its core foundation. As noted, GIS provides technology and methods for data integration, spatial analysis, and collaboration. GIS also provides a science-based framework for organizing workflows that integrate many factors needed in decision making. GIS improves the way we do our work by facilitating better decision making (saving money, time, and resources) and allowing us to communicate more effectively through geospatial visualization. In other words, GIS can help energy organizations better utilize their limited resources to meet business objectives. This includes discovering areas of weakness in the infrastructure that only spatial analysis could discover; for example, a GIS can provide a map that merges outstanding maintenance work orders (such as gas leaks or transmission tower corrosion) with crucial data (such as increasing traffic patterns or gas usage, heightened construction activity, or nearby excavation activity). A GIS can examine where a natural gas pipeline crosses a critical electric transmission corridor and then assess the risk in a way that most people can understand.

Well before a problem hits, a GIS-based dashboard can display key performance indicators of the business, showing where there are successes and where there are opportunities for improvement.

Energy companies around the world are applying GIS to virtually all the problems they face: site location for oil wells, gas gathering system asset management, environmental assessment, construction coordination, outage management, risk and integrity management, and mobile dispatch for maintenance and inspection. The sheer number of applications suggests that GIS is becoming a major instrument for understanding how we effectively manage and evolve these industries.

28.2.4 Problem and GIS Solution Patterns

While there are many problems facing utilities and energy companies, there are four problems a GIS platform is well suited to solving:

- Data management—for example, managing assets such as sensors, poles, conduits, smart meters, trucks, people, and imagery
- Analysis—for example, determining the optimal placement of fault indicators based on lightning strikes, or the next big wind farm
- Mobility—for example, collecting data from the field and integrating it with corporate dashboards
- Operational awareness—for example, visualizing the business spatially, such as identifying customers with small houses and high energy usage.

Risk Determination

A common problem facing energy companies is the ability, or inability, to manage information. For example, an insurance company establishes a price for a fire insurance policy based on an understanding of risk. If there was no information on the building's fire suppression system, the insurance premium would be higher regardless of whether the building actually had a fire suppression system. Information, or a lack thereof, determines perceived risk. Similarly, if a utility wants to build a very expensive high-voltage transmission line, it must decide whether the towers will be located in low-risk areas. Without information about fire potential, earthquake zones, vocal neighbors, vegetated wetland restrictions, flood plains, or soil conditions, the perceived risk becomes higher. With more solid information, companies are better able to assess risk.

Just as not enough information can be harmful, too much conflicting information can also increase risk. For example, investor-owned utilities pay municipal property taxes on their distribution assets. However, most utilities will admit that plant accounting property records do not exactly match engineering asset records. So, which count of poles does the utility use to calculate property taxes? Since they do not know which source is correct, they make an arbitrary choice. They choose either the source with the higher number of assets and pay more taxes than they need to, or the source with the lower number of assets and risk having an audit and owing money to the municipality. Both choices involve a higher risk than simply taking an accurate count of facilities. A single, reliable source of asset information available to both accounting and engineering, with a workflow that provides quality control, is cheaper to operate and carries a much lower risk (even if the accuracy is not 100%).

Regulatory risk is increased by inaccurate, inconsistent, or missing information. Most of the world's electric and gas utilities are highly regulated. Oil and gas companies are less regulated but must meet very stringent environmental and safety regulations.

For example, the President of the United States signed the Pipeline Safety and Improvement Act over a decade ago in 2002. The law outlined strict standards for oil and gas pipeline safety protocols. At the heart of the act was the need for pipeline operators to do a baseline assessment of the risk that their pipelines would fail. Depending on the level of perceived risk of explosion or leakage, operators are required to perform increasingly onerous and expensive pipeline inspections. The law defines a high-consequence area as one in which a pipeline rupture may cause a major problem, perhaps resulting in death and property destruction. Inspections could involve excavations or a shutdown of the pipeline. Risk was determined by information from pipeline operators. A mistake in the documentation or data-capture workflow could be hugely expensive. For many pipeline companies, information about the locations and conditions of pipelines is stored in various files, records, and notebooks. In audits of pipeline accidents, regulators have discovered inaccurate, out-of-date, missing, and conflicting information. As a result, pipeline operators were unable to make the best decisions. In general, pipeline risk factors increase due to a lack of good data management processes and/or inadequate workflows. Without accurate pipeline data, an operator would have to assume that things are fine and forgo a full assessment or assume that things are worse than he thinks and throw money at the problems. In both cases the operator is vulnerable. It is not reality that quantifies risk, but the quality of the information. Pipeline safety regulations require pipeline operators to prove that they have a quality control process for gathering, maintaining, and reporting information. The most successful strategy that energy companies and utilities have found is to collect, maintain, and report asset information on a map, driven by a consistent data source such as a GIS. In fact, the US government stipulates that pipeline information should be delivered to the Department of Transportation in shapefile format. Shapefile format is the de facto standard GIS data transport mechanism developed by Esri.

Four Common GIS Patterns

Figure 28.1 describes four common issues at utility and energy companies that GIS solves. They are data management, planning and analysis, workforce optimization, and operational Awareness. The diagram displays the challenge, then below it, the value of what GIS provides.

Data Management: Is the Information Accurate?

Energy companies are old. Some utilities have been operating for more than 120 years. They have built information systems using manual processes. To manage the workload, companies created many independent departments, such as engineering and plant accounting. The bureaucratic layers often inhibited the free flow of information. As these departments moved toward automation, they often created standalone departmental systems with no mechanism to verify information consistency. While energy companies have improved and modernized their information systems, many problems persist.

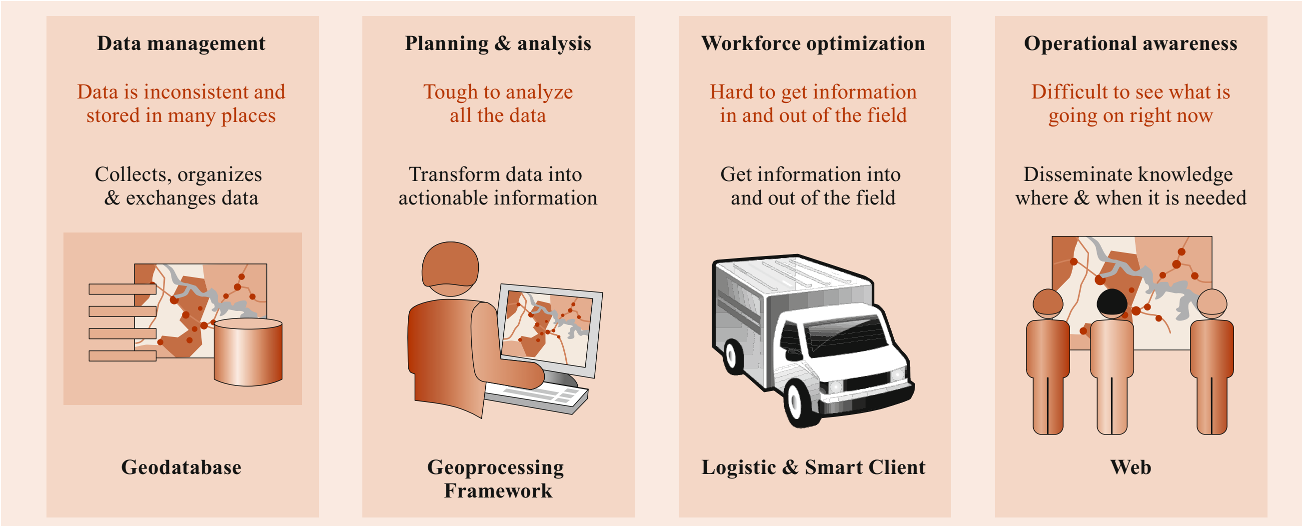


Fig. 28.1 GIS patterns

Asset information for energy companies is often incomplete, inaccurate, or inconsistent. Much of the asset management information is not easily verified. Most pipelines and many electric facilities are underground, and oil well equipment is sometimes buried deep within the Earth. Those facilities that are relatively easy to verify are widely dispersed throughout the service territory.

What is the best way to address these challenges? GIS is recognized for its strong role in managing traditional electric and gas transmission and distribution networks, pipelines, production facilities, gathering systems, and wells. GIS provides the most comprehensive inventory of these network components and their spatial locations. In short, GIS becomes utterly critical. GIS not only manages the data itself, but it manages data transactions. A solid quality assurance process can make sure that data is up to date and accurate. One of the key aspects of enterprise GIS is the workflows designed to ensure data accuracy.

Energy companies and utilities can rely on GIS to resolve problems that stem from poor data management. A GIS, by definition, is not a drafting system or a computer-aided design (CAD) system. Rather, it is a database management system that responds to queries by providing results in the form of a map. When an operator asks a GIS to *show all the gas leaks that have not been repaired in 2 years and that occurred on roads that are scheduled to be repaved*, the GIS responds with a map that highlights the appropriate pipe sections. The GIS platform facilitates the sharing of knowledge widely.

As discussed previously, GIS is not about creating a cleaner, clearer map from an old hand-drawn map. It is about discovering something new. That discovery should result in decisive action. In the previous simple example, the gas company can schedule the replacement of leaky gas pipes just prior to the city repaving the streets. Too often the opposite happens: the city repaves a street, and within months the electric or gas companies have to dig up the street to do a repair. The city then has to patch a newly repaved street. This common occurrence is both wasteful and inconvenient to the public. These events erode the credibility of the city and the utility. Furthermore, such mistakes make it more difficult for a utility to get approval for facility siting or a rate increase. Public displays of poor planning and misinformation can seriously erode a company's ability to do its job.

Planning and Analysis: What Is the Data Telling Me?

Even with good information, companies do not always have a methodical process of analyzing the data. With inaccurate, outdated, and inconsistent data, analysis becomes impossible. Many energy companies and utilities compensate by building workaround workflows. For example, a utility should know how to determine the proper location to install a gas regulator station. Without accurate data, the company will do analysis in the office and validate by taking field measurements. Utilities create job classifications to ensure that the information on the record accurately reflects the information in the field. An underground inspector, for example, checks a proposed cable route with field-based analysis—a physical check of available empty duct positions. The inspector has to climb into every manhole to see whether the empty conduit specified by the designer is still empty. If it is not empty, the inspector has to find an alternate route. This course of action might take several weeks to complete. The workflow is expensive and could be replaced with an analytical process that uses a routing algorithm and accurate data—a process GIS can perform in seconds.

Since such complex problems have spatial components, companies may not have the proper analytical tools they need for guidance. For example, an electric company has to make a decision about how to prepare for an impending storm. It knows the distribution system has strengths and weaknesses. It knows that some tree trimming was performed recently in some areas and not in others. It knows that some of the equipment is old and some is new. It knows where there are outstanding maintenance work orders. It also knows the history of failures for some of that equipment. The problem is that the utility manager must decide how many crews to deploy to which areas in order to minimize customer outage in the most cost-effective way. The utility manager must make a quick decision based on the information at hand. With spatial analysis, the manager could rely on more than disparate information and gut instinct. If all the information could be assessed together, the manager would discover the most vulnerable areas given all the known factors.

The power of GIS is that it helps the energy company understand how its assets are related to each other and to the surrounding environment. The ability of the GIS to perform complex analysis mitigates some of the risk arising from retiring employees. These older employees know a lot about the system's condition. When a thunderstorm is on the horizon, seasoned employees will instinctively know which parts of the system are more vulnerable than other parts. They carry utility data in their heads, having seen the condition of the system with their own eyes. When these workers retire, all of that mental analysis goes with them. While GIS cannot replace human intuition, it can provide the tools to qualify what people know to be true. A GIS can consume web services such as predictive weather. It can determine from inspection information which facilities have been maintained and which have not. A GIS can tell you whether tree trimming is current. It can capture information about infrastructure age and material. In the end, the GIS can provide a vulnerability assessment that shows exactly where the distribution system is most vulnerable, and where the storm will hit. This spatial analysis can then be used to stage crews and minimize customer impact.

As things get more complicated, GIS becomes even more important. For example, the smart grid is composed of two networks—electric and communications. Utilities must understand the physical and spatial relationships among all network components. These relationships will form the basis for some of the smart grid's advanced decision making. The smart grid must have a solid understanding of the connectivity of both networks. GIS provides the tools and workflows for network modeling and advanced tracing. GIS is used to determine optimal locations for smart grid components. During the rollout of the smart grid, utilities will need significant analysis to determine the right location for sensors, communication marshalling cabinets, and a host of other devices such as fiber optics in conduit and on poles. Since optimal device locations depend so heavily on the existing infrastructure, utilities will rely on GIS to support design services. GIS can provide a spatial context to the analytics and metrics of the smart grid. With GIS, utilities can track the metrics over time and provide a convenient means of visualizing trends. Since the smart grid is supposed to be smart, it will need GIS to provide advanced performance analytics, track trends in equipment and customer behavior, and record key metrics.

Likewise, oil fields need to be smart. Oil companies have invested millions in wells. Seasoned oil company employees have a gut feeling about how to do maintenance and which areas to focus on to optimize production. As these employees leave, oil companies must retain their knowledge about asset management, production, and maintenance issues.

A lack of spatial analytical tools is solved by GIS, as long as the data are accurate. Good data management, coupled with solid analytics, gives utilities the ability to make decisions about all kinds of business workflows, from engineering to financial to legal to customer service.

A GIS can help you analyze, for example, the relationship between frequent power outages and customer satisfaction based on a recent survey. It is a matter of a simple GIS query: *Show me on a map where there have been power failures and show me also where the customer satisfaction surveys show lower than average customer satisfaction.* If the result of the query shows a spatial correlation, you can deduce that lower customer satisfaction numbers correlate with higher outages. However, if you further analyze places where meters are not read or where bills are high, you might gain further insight into customer dissatisfaction. More factors, such as utility construction, could be added to the analysis. Without spatial analysis, a utility could come to the wrong conclusion as to why customers are dissatisfied. By performing a spatial analysis, things become clear. The result could show that customers in a particular area rated the utility lower because their calls received a higher number of busy signals. Utilities may make decisions based on false assumptions and thereby ignore the real cause of customer frustration. GIS performs these necessary and complex analyses, and presents results in the form a map that enables better decision making.

Workforce Optimization: Where Are the Assets?

Energy companies are complex. They have millions of pieces of equipment scattered over thousands of square miles of cities, towns, rural areas, remote fields, and mountains. Assets are strung on poles, buried in the ground, hidden in basements, and hanging high in the sky from huge towers. Usually, the vast majority of employees and contractors are deployed to field locations. Utilities have line workers, meter readers, and troubleshooters. Oil companies have field maintenance workers, technicians, diggers, and drivers. Energy companies and utilities have much of their hard and human assets scattered in thousands of different directions. This dispersion of workers and assets is a challenge. Companies need to know where the workers are in relation to the assets. When a transformer catches fire, it is critical to know which assets are in close proximity to the transformer and where the nearest troubleshooter is to the burning transformer.

The key to a successful field operation is solid knowledge of field worker location and activity. Of course, this is what GIS does best. Having access to mobile GIS in the field allows field crews to understand the location and the attributes of the assets. For example, GIS would tell a field worker the last time a device was maintained, its rating, age, condition, manufacturer, failure history, and any other critical piece of information. GIS will tell the field worker about buried or hard-to-access equipment. It will tell them about land, access locations, the surrounding area, sensitive habitat, and any incidence of crime or fire risk.

Likewise, field crews can provide information to the GIS, and to the office through the GIS. A field worker can update inaccurate data, bolster files with pictures of asset damage, and remotely record inspection information. Organizations with many field workers and assets often have difficulty getting information from the field to the office. In a utility, for example, field workers will typically capture as-built information on paper field sketches. These sketches then have to be sent to the field office or corporate office. Since the sketches are manually prepared in a somewhat hostile environment, they may not be clear enough for someone in a drafting office to understand or interpret. The time lag and the lack of clarity often results in a loss of accuracy and lack of timeliness of corporate data. Some utilities reportedly have field sketches that are more than 1 year old and have not been incorporated into the corporate data.

GIS helps utilities manage data about the condition of assets. After parts of the system go into service, utilities must maintain the system through the collection and maintenance of asset condition data. Some condition data can come from automated systems and others from inspection systems. Energy companies and utilities are rapidly adopting GIS-based mobile devices for inspection and maintenance. Enterprise GIS, with its desktop, server, and mobile components, allows utilities to gather condition data.

It is difficult to manage people in the field. A medium-sized distribution company serving one million customers might have a thousand or more employees deployed to the field. Just keeping track of that many people and vehicles is an enormous task. The GIS (along with other corporate systems) helps route and organize the field force for better utilization. Large infrastructure projects require enormous coordination for field forces and material flows. GIS helps energy companies see what cannot be seen by the field workers. It helps to visualize field coordination, route crews, see where material is stored and needs to be moved, captures field conditions, and fosters communication between office and field.

Operational Awareness: What Is the Data Telling Me?
Operational awareness is the ability to take data from a variety of sources and paint a coherent picture of what is going on right now. Control systems such as supervisory control and data acquisition (SCADA) systems only tell part of the story. A SCADA system, for example, only shows the electrical network in schematic form, and only shows information about the electric system. It cannot show a wildfire boundary in relation to a transmission line or an oil well. GIS-based situational awareness takes information from SCADA, work management, environmental overlays, and any relevant spatial information to display data in a single picture. GIS also displays results from spatial analysis such as vulnerability and data from a mobile device. This includes information from external web services—traffic, vehicle routing, etc.

GIS allows energy companies and utilities to visualize what they own and manage. If utilities have adopted smart grid or advanced metering infrastructure (AMI), or even extensive distribution automation equipment, their communications systems and electric facilities have to be managed. Oil companies need to see which wells are producing, where maintenance personnel are working, and where there are hazards or opportunities for improvement. Managers need to know where they are making or losing money so they can decide what to do next. Visualization with GIS goes well beyond the traditional *stare and compare* method commonly used by energy companies. GIS is about seeing relationships. With GIS you can monitor and express the health of the system in an obvious way with commands such as *show me all the sensors that have failed to report results in the last hour*. GIS can show a real-time view of the grid and note where things are changing. In effect, GIS (as compared with a SCADA system) shows the complete state of the grid, represented by a realistic model in a way that people understand.

Executives to managers to supervisors need to know what is going on right now. It is tough to see relationships when they are shown in lists. It is also too late to do something about a problem in the field if the reports are days or weeks old. GIS provides the spatial context to display analysis results, asset location, and field crew position. This quickly reveals actions that are needed now. Customers and senior management demand this.

28.2.5 GIS Helps Energy Companies Transform

People commonly speak of digital transformation. Too often, they confuse digital transition with digital transformation. Digital transition involves applying digital technology to legacy work flows. One common example is the migration of hand-drawn utility operating maps to a GIS. Yet they continue to print out those maps for field workers. An example of the difference between digital transition and digital transformation has occurred in the music industry. Since the beginning of the industry, music media has been analog—wax cylinders, plastic records and tapes. The migration from records or cassette tapes to compact disks was really a digital transition. People continued to buy and store musical content in much the same way as they did with old record albums or cassette tales. Today, with streaming music services, the entire experience of dealing with music has been transformed. While technology is a component of digital transformation, the difference is in the behavior of the users.

A GIS enabled through platform technology can create a digital transformation if energy companies do not simply apply the technology to legacy work flows and processes. Sure, true behavior must change. GIS has the capability to answer tough questions such as *where in my infrastructure could a single event could bring down the pipeline, refinery, or electric network?* GIS can provide the backbone for a self-healing grid. It can tell gas operators where they have the greatest risk of violating safety regulations. It can help utilities and energy companies discover new things about their investments and risks. Unlike any other technology, GIS can perform spatial and network analysis and show the results in the form of a map. GIS is not just about mapping; it is about discovery. Getting that information to everyone without file transfer or paper reports and maps can be transformational.

To meet the significant challenges of today, utilities and energy companies will have to use every resource available. Many utilities that have built a GIS have a wonderful resource ready to meet their needs and challenges in the years ahead. GIS provides a model of the utility network and energy infrastructure. It captures the inventory, location, asset condition, sensor networks, smart meters, monitoring and control systems, and all the communications systems that link technology. Through real-time maps, GIS displays the relationships between those assets and other infrastructure and the surroundings. Through modeling, GIS provides a visualization of the impact of company decisions. GIS must be a fundamental tool in the deployment of a smart grid, a smart pipeline, a smart field, and all the *smart* systems that will be required over the next decade.

Governments are promulgating tough regulations that will dramatically change the landscape of the energy business in short order. In the USA, several states have banded together to form carbon cap-and-trade systems. Some states have established very tough renewable energy targets that will stretch the resources of utilities and energy companies. Society will challenge the coal industry to discover ways to sequester carbon dioxide from burned coal economically. We may well see the development of a brand new infrastructure entirely devoted to a carbon pipeline. As this book is being published, the US Congress is contemplating a federal carbon cap-and-trade system.

It is becoming clearer that the transportation industry will become more and more electrified in the years to come. Given that the world's electric infrastructure is old and overtaxed, this will be a further challenge. Oil companies are faced with depleting fields and will need to squeeze every ounce of oil from the existing infrastructure. GIS can help with the siting, assessment, and evaluation of all kinds of energy assets and infrastructures. The natural gas industry has uncovered new and innovative techniques to extract gas from the ground. Fracking has dramatically changed the face of the entire energy value change. Location plays a fundamental role in determining where to apply this technology and better understanding the potential risks.

28.3 System Design

28.3.1 The Structure of an Energy/Utility GIS

An easy way to understand an energy/utility GIS structure is to think of the data as being grouped into a series of layers. Each layer represents a theme, or set of features that have a strong relationship to each other. For clarity of purpose and ease of representation and modeling, it is helpful to group similar features together. For example, geological features that help represent underground energy resources such as coal, oil, and natural gas would logically be grouped together. Depending on the complexity of the layer, it may be prudent to create sublayers of more synergistic features. In a utility model, it makes sense to group land features together. Most energy companies and utilities structure their GIS in such practical ways.

- Subsurface layer: This layer includes underground features that help energy companies manage and discover resources. For a utility that either operates or is seeking geothermal resources, this same layer would be developed.
- Cadastral layer: This layer includes land features that have to do with ownership and subdivision. This layer holds survey data and information about parcel boundaries, street rights of way, utility easements, political boundaries, and environmental overlays such as wetland delineation.
- Street layer: This layer is related to the cadastral layer; however, the street is also a linear network with unique characteristics that enable vehicle routing and geocoding. Street layer data include paved width, address ranges, and intersections. The street layer is modeled as a series of lines (street segments) and nodes (intersections) connected together to form a linear network for the purpose of flow (finding the shortest route from an electrical substation to a customer location). It also serves to facilitate

the Global Navigation Satellite System (GNSS) location of a feature, such as a house, based on the address of that feature or the nearest address.

- Structure layer: This layer is unique to the type of company being modeled. It shows structures that support the electric and gas equipment for a utility or the support structures of an oil well. Features modeled within the structure layer include electric vaults, steel or wood poles, oil rigs, transmission towers, and regulator station structures.
- Facilities layer: This layer models wires, pipes, and facilities that carry the energy product (gas, oil, electricity). Examples of these kinds of features include electrical conductors, gas mains, valves, an oil gathering line, an electrical transformer, or a switch. There may be reasons to create sublayers for these features; for example, it might be helpful to separate the electric or gas transmission system from the distribution system.
- Customer layer: This layer models the locations of the end users. This includes commercial, industrial, and residential locations for utilities, and wholesale or retail locations for oil companies.
- Real-time layer: This layer varies widely depending on the industry and application, and may contain many sublayers. It is often the repository for web services such as vehicle location, weather and traffic services, or SCADA data.

28.3.2 Energy/Utility Data Models

A collection of data models provide the data management framework for an energy/utility GIS. Generally, there is a data class for each layer. The data class details each asset class (the particular asset within the layer), its attributes, behavior, and the relationship of the asset class to other asset classes. For example, the attributes of the GIS model for a simple utility pole contain certain valid values. A pole that is more than 30 m or less than 5 m is not possible, so the data model adds that behavior to this asset class. There is a relationship between a utility pole and an electrical wire. This relationship is modeled within the data model.

The data model represents a description of the information that will be stored in the GIS database. It describes the semantics and the extent of the data to be contained within the GIS database. While the data model itself does not describe a GIS application, it must be able to model all data needed for every foreseen application. For example, if a gas utility is going to use the GIS as a basis for its distribution integrity management program, it would have to include data, relationships, and behavior to define the integrity of the system and the risk associated with operating the infrastructure.

28.3.3 Cadastre and Land Management

Cadastre is a legal representation of the land, and it should be as close to survey accuracy as practical. It includes all important political and zoning boundaries, rights of way, easements, leases, and licenses (Chap. 20).

Other land information important to energy companies and utilities includes:

- Land features such as railroads, points of interest, monuments, and unusual land features
- Waterways and regulatory features in the service territory
- Indigenous burial grounds, sensitive habitat, and migratory patterns
- Real estate owned, maintained, leased, or licensed by the energy or utility company
- Outstanding operating permits
- Building footprints
- Imagery
- Street networks
- Digital elevation information
- Vegetation types.

The collection of land information features noted above is often referred to as the landbase or the base map. It is necessary to build attributes, behaviors, and relationships for all land features. For example, an attribute for a utility easement may include the date the easement was granted and a digital representation of the legal document as well as easement restrictions.

28.3.4 Energy Company Workflows and GIS

Figure 28.2 shows a typical workflow life cycle. Again, GIS plays a vital role in each, as detailed below. An energy com-

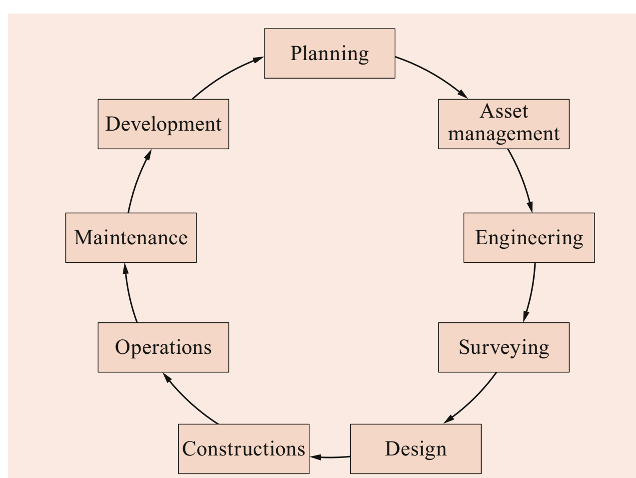


Fig. 28.2 Energy and utility workflows

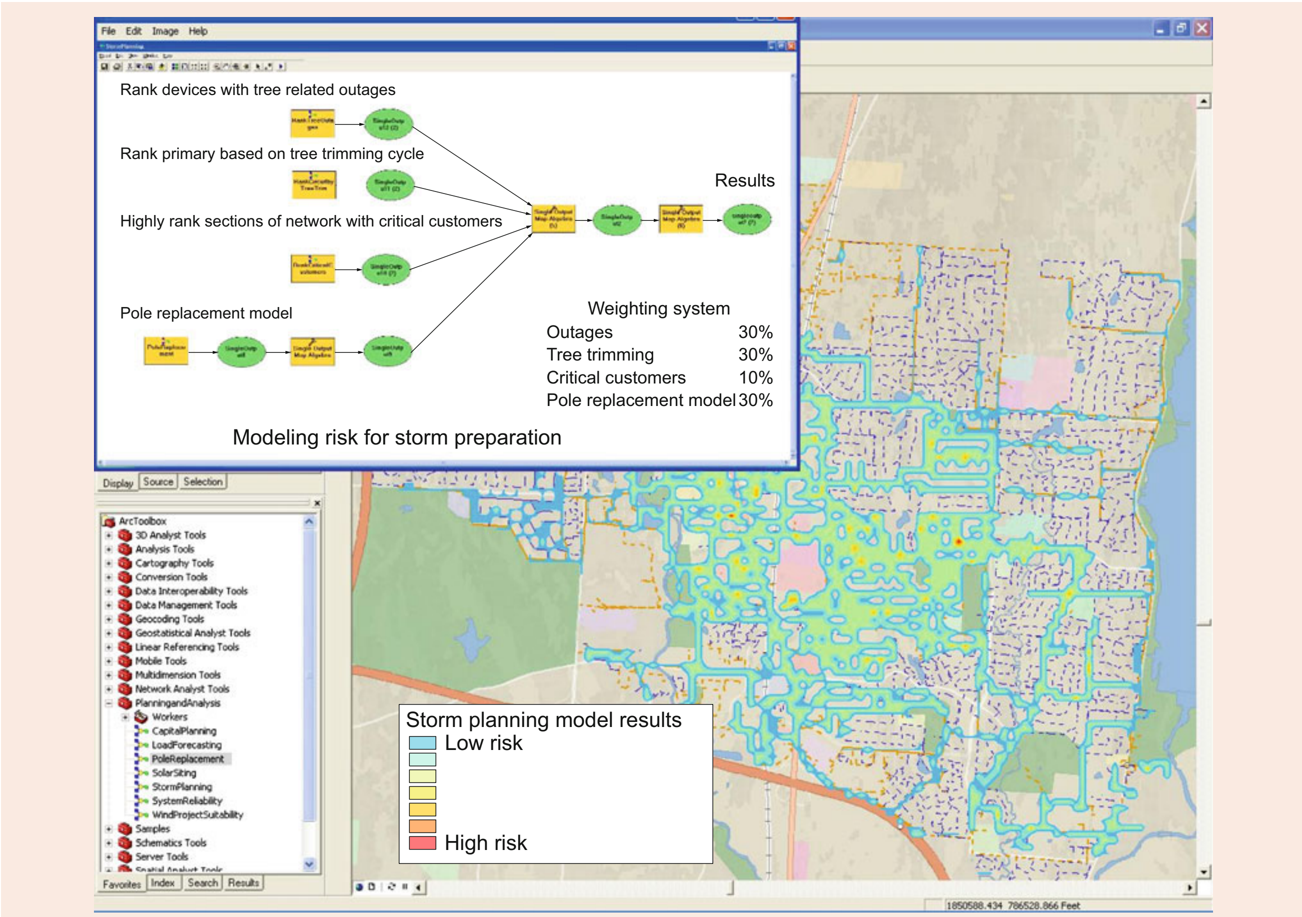


Fig. 28.3 Planning and analysis

pany's infrastructure is constantly changing—adding lines, installing new oil wells, adjusting gathering systems, aligning pipelines, connecting customers, and plugging leaks. Engineers plan for increases in capacity for a new part of the city. They replace cast-iron mains with plastic mains. Like most enterprises that operate and maintain complex infrastructures, these companies follow a particular work lifecycle. They have very specific workflows that tend be cyclical and follow a definite pattern. GIS often plays a role in each of these workflows.

Planning

Figure 28.3 illustrates how GIS gives planners more detailed information about the area they are working in. For example, the chart shows a model that allows planners to identify various risk factors contributing to power failures. The model brings in, for example, heavily treed areas, the age of assets, and the location of critical customers, such as water pumping stations. The maps show areas of high to low risk based on the analytics provided by the model. Projects do not just happen. Planning workflows involves an understanding of where we are now and what needs to happen next. Utilities regularly perform load forecasting based on what they know now about current electrical consumption. They need to predict what the load will be in 1 day, 1 year, and many years into the future. They also need to determine what impact that load will have on the infrastructure's ability to deliver its product. GIS manages the current data. It gives utilities a way of quantifying new demands. GIS can analyze data such as land use, zoning, and demographics to predict future demand size and location accurately. This data can then be weighed against the infrastructure's ability to meet impending demands. Gathering data is common and often critical for planners. Once the data are collected, a project can proceed. The four patterns of GIS—data management, analysis, data capture, and situational awareness—are used fully in the planning workflow.

Asset Management

The asset management process also yields projects. Rather than predicting the future demands of the system, asset management predicts the future condition of the system and determines what to do about that knowledge. Properly done, the asset management workflows require a solid inventory of the assets, including asset location, age, size, rating, condition, and relationships to other assets and communities. GIS captures assets and references them in the real world. GIS has the ability to capture, or at least reference, asset attributes and condition, and to determine the assets' relationships to other things. However, GIS is not just about showing information on a map; it is about discovering new information that is actionable. For example, a GIS can show which part of the gathering system's repair costs have exceeded the cost of a new pipe section. It can show vulnerability if leaks are in hard-to-access areas, or if there is a combination of age, frequent maintenance, and higher demand. The asset management process could (and often does) result in major maintenance as well as new capital projects. Without GIS, asset managers would need to rely on tabular reports, paper documents, and the memories of the workforce to come up with a defendable and science-based analysis. The four GIS patterns are strongly used during this workflow.

Engineering

Once a project is approved, or at least under serious consideration, the engineering workflow begins. In many ways, the engineering workflow has elements of both planning and asset management. Once the system planners have determined a need for a new transmission line, engineering can begin. This process takes the technical details and balances cost, risk, and reliability to come up with several alternatives that meet the needs as defined by the asset managers and planners. Engineers select various materials and technologies and ultimately make a decision about where the new project needs to go. They often base these engineering decisions on location, land features, geopolitical and environmental risk, and cost factors. Engineers need good field data, existing condition information, and external information from sources such as conservation commissions and registries of planning and environmental groups. GIS allows engineers to mash up or collocate spatial information from different sources, providing new insight and analytical tools to quantify risk. With GIS, an engineer can determine and map optimal areas to route a transmission line in a way that is easily understood and acted upon.

Surveying and Rights Procurement

Once a project is initiated, the company will need to exactly locate the property or properties on which they plan to put their new assets. For example, wind farms are becoming common. However, as stated previously, there is often community opposition. Consequently, it is critical that wind farm developers have a clear understanding of their land rights to the property they propose to use and the spatial planning legislation framework, otherwise the entire project could be put at risk. Even routine projects need careful land ownership information. GIS has become the tool of choice for surveyors as well as planners, asset managers, and engineers. In the past, survey systems did not really communicate well with GIS. Today, those technical issues have largely been resolved. Since utilities need public access to land that they do not own, they must have a clear understanding of the limits of their rights. Many municipalities require energy and utilities to prepare petitioning plans to grant rights for placing poles, towers, mains, and any other equipment installed in public areas. Since granting location has to be accurate, utilities hire surveyors to create the plans. Utilities submit these plans for approval by the public works commission or municipal council. Once approved, the plans become legal documents granting location. If the utility needs equipment on private property, it has to obtain licenses or easements. The utility must have precise surveys for the legal description of the easement. Design engineers may require test borings to determine if there are underground obstructions that are not marked on reference documents. Surveyors need to capture the locations of any obstructions. GIS becomes an integral part of the success of the surveying and rights procurement workflow.

Design

The design workflow utilizes the work done by the planners, asset managers, surveyors, and field researchers. This collection of work is spatial and can easily be managed within GIS. Designers can detail the project so it can be built. Here, the strong data management and spatial analytic capabilities of GIS are crucial. Designers must lay out lines, access points, poles, wells, and gathering structures that span several feet to hundreds of miles. They need to account for the material, the lengths, the quantities, and any special conditions. As designers lay out the project, they need to be aware of potential interferences, obstacles, and clearances. These can all be managed within the GIS. Many utilities are using GIS—often with third-party extensions or specific applications—to include direct design and construction documentation. This is typical for electric and gas distribution, less so for electric and gas transmission, and rare for substations, meter stations, and refineries. However, GIS provides much-needed spatial analysis and data management to the design workflow.

Construction

The construction workflow is the most intensive and expensive workflow for energy companies, yet this area is where GIS is least likely to be used. Construction involves more than hammering nails and welding steel. It can be a logistical nightmare. Often schedules are tight, budgets are constrained, and work involves the disruption of the community, including traffic patterns. GIS can be effectively used to manage the volume of information coming into the construction manager's office, from field designs and work order changes to material storage issues. During the construction phase, the four patterns of situational awareness are perhaps the most dominant part of the workflow from a GIS perspective. Whether the construction workflow is the building of a 10 mile pipeline or the simple installation of an overhead service to a single-phase customer, location is almost always involved. When multiple things happen in the same geographic area, communication and collaboration can suffer. GIS can help companies visualize situational awareness. Documents, deliveries, spills, accidents, change orders, crews, security guards, traffic reroutes, rubbish disposal, and hazardous-waste transport are only a few considerations that make GIS critical during construction. As noted earlier, better information reduces risk. Additional risk on a construction project always translates into additional costs. Clear, consistent geographic knowledge lowers risk and thereby lowers the cost of the project.

During the construction phase, the use of platform technology can save money and time. Rather than relying on conventional reports and paper field sketches, a GIS platform can use the three c's to help communicate to all parties, collaborate with vendors and workers, and coordinate the many doing the work.

Construction Documentation and Reconciliation (As-Builts)

Without GIS, there is a gap in the natural flow of data from planning to design to construction. When all updates are made to the GIS during construction, the as-built and construction documentation processes are more efficient. When this is not the case, two things happen: first, the as-built information never gets captured in a central GIS accessible to all, and second, the information represents a large volume of work that sits in a pile on someone's desk, waiting to be processed. In both cases, the information that operators use for day-to-day decisions is flawed. The as-built workflow, or construction reconciliation and documentation workflow, is a workaround workflow that compensates for a lack of consistent data management and field data collection during construction.

The GIS platform closes the gap between when something happens in the field to documenting that event for all to access.

Infrastructure Operations

After the assets are installed, the buildings are occupied, the transmission lines are energized, the windmill is spinning, and the oil and gas are flowing, energy companies and utilities must operate the business. Clearly this requires a solid understanding of where things are located, their condition from an operational perspective, and what is happening now that may impact their operations. From a data management perspective, this often involves mashing or combining data from other sources, such as a real-time control system, a weather service, or the human resources system. GIS provides the operational view to make decisions, which are almost always spatial in nature. *Where do I deploy extra crews during a storm? Which part of my gas pipeline system is vulnerable to terrorist attack? Where are my crews right now? How much money am I spending right now? Am I spending that money on the right things?* GIS provides spatial context and data management, along with analytic tools, mobility, and operational awareness. With GIS, utilities and energy companies move from gut-instinct decision making to process-driven, defendable actions based on accurate, up-to-date information and rigorous analysis.

Maintenance

The quality and thoroughness of maintenance, and its related inspection workflow, will separate a mediocre energy or utility operator from an outstanding one. Much has been written about reliability-centered maintenance or the workflows around optimizing maintenance. These are powerful and necessary things. However, spatial context provided by GIS enables two important factors:

- The ability to perform complex spatial queries on infrastructure, maintenance, and inspection data within the maintenance management system
- The ability to visualize the impact of an increase or decrease in maintenance and inspection.

GIS can also be used for the very tactical day-to-day process of managing the maintenance of a vastly geographically distributed set of assets. When customers are added as assets, the sheer number of things that energy companies have to worry about is momentous. As utilities roll out smart grid equipment, the complexity and quantity of assets in the field will increase along with the need for a strong GIS platform to support accurate data and effective data maintenance systems. With GIS, utilities can better understand the patterns of failing infrastructure by analyzing probable causes of problems with consideration for geographic location. In addition, GIS supports day-to-day inspection workflows, such as laying out an optimal inspection route based on the actual parameters of traffic flow, total distance, timing of stops, and time of day. GIS factors in all the large and small issues to ultimately help companies make and keep more money.

28.3.5 GIS and the Energy Company Value Chain Master Processes

As noted, energy companies are complex organizations. During the twentieth century, most electric companies were vertically integrated. Some utilities owned and operated coal mines and the short freight lines that transported the coal to the power plants. Today that is much less common. Some energy companies do have business units that cover exploration and retail delivery of gas.

What follows is a discussion of the energy value chain regardless of whether the chain is operated by one company or many. Within each major element of the chain, all the workflows typically come into play. For example, within the electric production division of a major utility, we will see planning, engineering, surveying, design, construction, as-built, operations, and maintenance. The same is true for an oil well development or a project that results in a new refinery. The energy value chain consists of the master processes that result in the delivery of energy from its source to the consumer. What also follows is how GIS has a role in each link in the value chain.

The energy value chain looks like this:

- Exploration: This master process has all the geological processes, data, people, equipment, tools, and knowhow to discover sources of energy. This is about finding oil, natural gas, wind, solar, geothermal, tidal, uranium, or any other source of energy that will be developed, transported, sold, and used. What kicks off this series of processes is the request or need to find a new source of energy. The end result is the actual discovery and verification of the source of energy.
- Development: Once the energy source is discovered, companies focus on acquiring that energy. An example would be the development of a series of single windmills or oil wells, and all the associated equipment.
- Gathering: Some companies run gathering systems. This critical link takes all the point sources of energy and gathers them into a single source to be refined or produced. An example would be a series of tractors and trailers gathering up cellulose fiber and dumping in a big bin to be used for biofuel. Gathering could come from a single field, like an oil field, or from many disparate sources.
- Production: Gathered energy is refined or generated from its sources and is readied for wholesale transportation. One could think of this link as where the manufacturing of the energy is performed. A nuclear power plant is an example of a production facility. An owner of a nuclear power plant may not own any other part of the energy value chain, but will likely exercise all the workflows and use all four GIS patterns.
- Bulk transportation or transmission: The input to this link is the finished product (or reasonably so) and is ready for transportation. Companies that transport include pipeline operators, electric transmission departments of utilities, and some standalone transmission companies. There is a high degree of control and coordination that exists between production and transportation.
- Distribution: This is the more local transport of the energy, which includes the trucks that deliver gas to gas stations and the local natural gas distribution company.
- Retail: People in this business will tap off the distribution system and sell directly to the end-use consumer. An example is a local gas station. This could also be the local utility who bills the customer and owns and manages the meter.
- Consumer: They use the product and pay the bill.

The sections that follow discuss how GIS is used in each of these major links in the energy value chain. While the industries are somewhat different, they all share common applications.

28.4 Applications

Historically, applications have been complicated and often multipurposed. Platform technology has changed that paradigm. Today, applications are smaller and focused. Often applications are simply configurations of repeatable workflows. For example, a simple damage assessment app may be a reconfiguration of an inspection app. Large multistep applications can be broken into smaller, simpler, and easy-to-use apps. The following are common GIS applications. Most are based on some common themes of data management, mobility, analytics, and situational awareness.

28.4.1 GIS for Exploration

Whether a company wants to discover a brand-new source of energy or to figure out how to get more energy from an existing source, the inputs and outcomes or outputs are generally the same. They all start with a request of some kind, based on an exploration analysis; a regional or global demand; a need to make more money; a shortage; or a government grant. It could also start with a farmer thinking about planting corn or cellulose to be ultimately used for biofuel production, or a coal mining company could be looking for a new vein or a lower-sulfur coal source. The process involves planning for the discovery, data gathering, engineering, land management, design of a testing system or drilling rig, construction, operations, and eventually maintenance. GIS is present during the entire master process, from data gathering to complex spatial analysis to resource discovery and energy development. Figure 28.4 details the various processes used by energy companies, from finding the resources, for example, natural gas, to the end-use customer's consumption of the resource and the various steps in between.

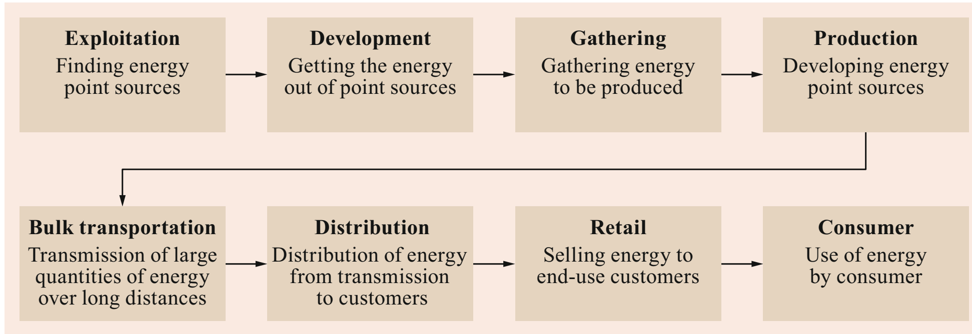


Fig. 28.4 Energy value chain

In the exploration master process, GIS is typically used for the following:

- Mapping potential sites. GIS can manage satellite imagery, digital aerial photo mosaics, surficial and subsurface geology, cross-section interpretations and images, spatial extent of a known resource, seismic surveys, well locations, the cadastre (land ownership, leases), existing infrastructure information, and the surficial and environmental elements that warrant consideration.
- Specifying the locations of field surveys. GIS is used to position seismic surveys, extract core samples, drill exploratory wells or digs, and determine where to place logging equipment or field measuring devices to help in the discovery process.
- Collection of field measurements. Commonly performed with handheld GIS mobile devices, data collection can provide information on land features that is otherwise not shown on existing maps. It also captures surface geology, vegetation condition, soil condition, water tables, wind density, and proximity to people, transportation networks, workers, resources, and markets.
- Collecting information from external sources. As a prelude to the analysis phase, data must be gathered from a variety of sources. This increases the chance of project success and lowers the overall cost. These data are often in various forms and projections. GIS helps to normalize and coordinate disparate data.
- Analyzing the data. Before any drilling, planting, or building, companies must perform comprehensive analysis integrating the relevant factors. Along with rigorous algorithms, GIS provides the spatial component along with visualization of the analysis results. For example, a GIS-based play (a group of petroleum fields/prospective in the same region or created by the same geological circumstances) risk analysis process is used to consistently and systematically value a play based upon chance of success.
- Selecting the most probable sites. Additional analysis based on nonscientific results relates to more practical project-related decisions, for example, determining which mining site to develop first, or how best to access the known resource from a variety of locations.
- Analyzing the suitability of the sites. The final results of all the exploration work are presented as the output of GIS. At a petroleum company, this may be the presentation a geologist makes to senior management to secure funds for development. Alternatively, this may be the materials used to convince a bank to supply the development capital for a corn biofuel development.
- Equipment and resource management. The entire exploratory cycle requires expensive equipment and skilled employees. Energy companies use GIS to route resources effectively and thereby get the most value from their equipment and resources.

Considerations for Geothermal Plants

The exploration master process for a possible geothermal energy source is an interesting example. Geothermal looks attractive on paper because its operation characteristics resemble those of a base-load coal plant. Unlike solar and wind power, geothermal power production is continuous. A geothermal plant leverages the heat below the surface of the Earth. This heat energy is transferred to the surface and used for movement (e.g., driving a turbine) to create electricity. There are no emissions and no fuel. There are, however, some environmental dangers to be considered. Since the energy from a geothermal plant comes from deep within the Earth, there must be a heat transfer process. Some detractors are concerned that the heat transfer medium (saltwater or some chemical) could contaminate drinking water aquifers. GIS can model those aquifers and support scientific studies to either confirm or quantify the risk. In addition to the environmental services of GIS, it is an essential tool to guide developers to the optimum location for a geothermal plant. Drilling is expensive, so the deeper one has to drill, the more expensive the project. Also, the resources must be high quality to justify the expense of the drilling and the infrastructure involved. Some geothermal resources are weaker than others, e.g., if the heat is transient. The effectiveness of the plant depends on the underground structures that permit heat to flow to the surface. In some cases, there are clues, such as geysers, that tell explorers that structures are nearby. So, mapping this activity in GIS can provide some insight into possible development sites.

In addition to finding the right geology for a geothermal plant, the spatial assessment takes into account a multitude of factors, such as groundwater protection issues and proximity to transmission, transportation, and labor sources. Since geothermal plants are often near tourist attractions, such as geysers, they are difficult to permit. GIS can be an ideal tool for developers to use to help the permitting process.

Siting of Wind and Solar Energy Sources

Wind and solar power is variable and therefore not as consistent an energy source as coal, oil, gas, or nuclear baseload power plants. While viable, these resources present challenges. Wind turbines can harm birds and can be loud and visually obtrusive. To justify the cost of a wind farm, the atmospheric conditions must be strong and constant. Wind farm planners use GIS to locate the most favorable source of wind energy. Planners also need to incorporate bird habitat and migratory patterns into their analysis. A viable wind farm needs wide expanses of land. Planning a wind farm, or even a single wind turbine, is a complex spatial problem that can benefit from GIS. Solar energy has the smallest environmental impact of all the common sources of energy. It is silent and completely clean. Finding the right environmental factors (such as in the desert) can be tricky. With GIS, users can retrieve the solar densities throughout the region at any given time. They can combine solar potential and energy demands to find the optimal and most economical location for solar energy development.

The challenges of locating and establishing wind and solar power production facilities can be overcome with GIS-based spatial analysis and decisions based on thorough consideration of data.

28.4.2 GIS for Energy Development

The next element of the master process in the energy company value chain is development. Once resources are discovered, they must be developed. A natural gas or oil well needs to be drilled; a reflective passive solar panel needs to be erected; a biofuel field needs to be planted.

Siting and Building an Oil Well or Mines

Locating a reserve is an expensive undertaking, and to recover these costs it must be put into production. To do this, the company must understand the geographic, infrastructural, and environmental factors as well as the business conditions in the region proposed for development. In the petroleum industry, technical considerations are used to screen and evaluate a reserve. Companies look at volume, structure, trap, reservoir, and charge combined with economic variables such as net present value and investment/profit ratio. In mining, technical considerations might shift to overburden, volume, grade, hydrology, etc. GIS modeling is ideally suited to this kind of overlay analysis. It allows a user to integrate complicated technical, business, risk, environmental, and economic factors and display the results in a simple, understandable way.

If a reserve proves to be economic, a production program is initiated. This program may require many wells to be drilled, an open pit to be planned, pipelines to be built, a slurry pond to be designed, or treatment facilities to be constructed. For oil and gas, only a certain number of wells can be placed in each section, and GIS is used to define the optimal pattern for a reserve. In mining, costs rise as overburden is removed or shafts are created, and GIS is used to help minimize expense. The technology also helps properly route drill or dig equipment and specialized personnel. GIS plays a role in selecting the best route for slurry, gathering, and transmission pipelines so that environmental and regional factors are mitigated while minimizing costs. As facilities are constructed, engineers use GIS to select a location and examine grading, runoff, vegetation, hydrology, environmental considerations, etc.

When developing any type of resource properly, all construction workflows play a role: planning, engineering, surveying, design, construction, and as-built. Once the resources go into production, the maintenance, operations, and asset management workflows continue. In each phase, GIS plays a key role.

28.4.3 GIS for Gathering Infrastructure

After energy sources are developed, companies begin the process of gathering. Point sources such as oil wells, wind energy turbines, solar panels, biofuel crops, and tidal resources must be pulled together into a production facility. The resources can then be transported, delivered, sold, and used. Again, all the workflows that use GIS are involved in the building and the operations of gathering systems. Wind farms consist of various sizes of wind turbines spread over rather large areas. Each wind turbine is a self-contained energy source consisting of large blades, a generator, controls, and the systems needed to monitor and manage the production of electricity. Electric cables carrying relatively low-voltage electricity are routed from each turbine to a central substation. Similarly, solar farms have separate arrays of solar panels or passive solar systems (mirrors, heating elements, and small turbines). The electrical system that connects these power and control cables through a series of underground duct banks, conduits, and cable trays is the gathering system. A gathering system needs GIS when it is planned, engineered, designed, surveyed, documented, operated, and maintained. The assets in this system need to be managed just like a conventional electric distribution system in a city or town. Other gathering systems include coal mining transportation that transfers coal from mines to trams and trains to coal cars on freight trains. A gathering system for biofuel involves tractors and trailers to gather and bundle crops to be processed.

Liquid and Natural Gas Gathering Systems

Liquid and natural gas gathering systems can be very extensive (they sometimes cover many miles), and may consist of great quantities of equipment and infrastructure. In all cases, the operator's goal is to keep the product inside the pipeline and minimize the cost associated with doing so. To do this effectively, they must know where their infrastructure is, its condition, and the system's operating parameters. Thus, GIS plays a fundamental role in asset maintenance programs and provides tools that work with existing operations workflows.

GIS is the repository for storing pipeline and equipment attribute information such as the material, diameter, age, condition, and maintenance/inspection records. With the modeling and mapping benefits of GIS, this information is used for planning maintenance programs, routing staff, optimizing production, assessing segment rupture risks, completing integrity management reports, and quickly creating accurate maps and detailed reports to support numerous other operations activities. A GIS captures all the vital data and centralizes it for easy access.

A gathering GIS system can be integrated with other enterprise systems such as SCADA or financials to provide a spatial view of the operation. Data about performance, real-time monitoring, work crews, new projects, and maintenance can be brought together from several systems into an enterprise GIS and communicated visually. Delivering the data geographically gives better insight into exactly what is going on with the gathering system, allowing operators to model operational, environmental, and financial risks.

28.4.4 GIS for Production Facilities

Production facilities include power plants, refineries, fuel processing plants, substations, and centers for collecting renewable energy point sources. These facilities take energy from gathering systems and create an energy product ready for wholesale or bulk transportation to distribution centers. In the case of a power plant, the production facility converts the energy from one medium to another, for example, the conversion of natural gas or coal to electricity.

A critical aspect of the master process of energy production is the selection of the site for the facility itself. Unlike the siting of energy sources, the siting of the production facility is less dependent on the location of the energy sources. Proximity to the energy source is a consideration in some cases but not always a requirement. Other considerations such as transportation amenities, public safety, and the environment may be of higher priority.

The site selection for an energy production facility is quite a different process from the siting of energy sources. GIS helps energy companies determine the best location for a large energy production facility (for example, a nuclear power plant) by examining the siting data and performing extensive spatial analysis. GIS helps companies take into account a number of spatial factors. Data are merged together (in layers, for example) to determine a series of possible locations. Results are prioritized using the spatial and visualization capabilities of GIS.

Considerations include:

- Land availability: GIS datasets contain population measures, land ownership, acceptable zoning overlays, and market values.
- Land suitability: GIS analyzes microclimatology for wind flow and direction, land topography for slopes, soil conditions, and hazards such as earthquakes or flooding.
- Environmental concerns: GIS datasets and analysis determine environmental issues, including spill impacts, pollution, habitat, and water availability.
- Initial construction: GIS can help analyze the cost of construction based on material availability, site transport, labor pools, existing infrastructure, pipeline or network accessibility, and local conditions.
- Operations and maintenance: GIS can be used to predict costs based on climatic conditions, waste disposal, access to labor, and social issues.
- Fuel transportation: GIS can analyze the cost of fuel transport.
- Access to employees: GIS performs demographic studies using business analytics.
- Community impact and opposition: GIS maps likely neighborhood issues such as noise pollution, odor, spill impact, and visual intrusion situations.
- Proximity to existing energy transportation systems: How close will the new power plant be to an existing transmission line? GIS does the analysis.
- Evacuation analysis: How easy it would be for employees to evacuate the site can be determined by GIS. How effectively can a critical region be evacuated in the case of a burst or spill?

GIS can consider all of these factors together, since they are ultimately spatial considerations, and optimize the location of the plant. Planners can combine the physical and social suitability in GIS. The result is a map that illustrates new knowledge about all the available sites to provide enhanced decision making.

Coal plants need to be near a ready source of coal, and of the kind that makes the plant effective. They need to be near trucks, trains, and boats. Nuclear plants must have an adequate roadway network to transport sensitive fuel, and cannot be too close to urban centers in case of an accident. Hydroelectric plants must be on waterways. Natural gas plants must be near pipelines. Oil refineries need to be near pipelines and transport mechanisms. Plant operations must have access to qualified people to work in the plants. Planners use GIS to locate plants close to necessary resources. Utilities seek locations where they can adequately protect the plant from vandalism or terrorism. Operators must consider if the proposed plant site is near high-crime areas. They have to be aware if there has been known terrorist activity near the proposed site. They also have to consider escape routes for their employees. In the case of an incident, catastrophic or minor, planners need to know the proximity of the proposed plant to the nearest police station or hospital.

28.4.5 GIS for Transportation (Transmission)

GIS for Transmission Pipelines

Regardless of the location or business focus, the goal of every transmission pipeline company is to transmit product from one location to another safely and cost-effectively. Working against this goal are the natural elements that consistently decay and erode a pipeline, and the human population, who damage the infrastructure. To maintain a current assessment of the integrity of their infrastructure, pipeline companies consistently collect information about the location and condition of their pipeline. Most pipeline companies have been collecting and storing these spatial records (construction, material, crossing, repair, right of way, etc.) for decades. Transmission pipeline operators have long benefited from GIS technology because the breadth of information has proven too significant to analyze effectively using tabular or database technologies.

Using GIS, pipeline operators integrate this information into a single spatial database and retrieve asset records as desired. Record integration also allows operators to answer complicated questions about integrity management, perform automated compliance activities, and generate regulatory submissions. Operators are able to optimize business with GIS through predictive analyses, production management, and preventative maintenance. In addition, this wide range of functions is performed without significant staffing changes.

For many operators, enterprise GIS has become the data hub used to address a multitude of business questions and organizational problems. Through GIS, information is shared across departments and integrated with other systems such as resource planning, work order management, supply chain, or document management. Anyone within the company can quickly search through massive amounts of data using geographic location or specific queries. Below are some of the departments and how they use GIS.

Routing and Engineering

Optimal planning of a large-scale pipeline development project requires the integration of information from a wide range of disciplines: engineering, environmental, socioeconomic, etc. Factors that are commonly involved include population settlement, available land, total pipeline length, land types, vegetation, surface geology, environmental considerations, and slope. GIS is used to compile this information, and an automated model is employed to combine the measurable features according to a predefined weighting system. This weighting system allows the best, or several appropriate, corridors to be selected. Using GIS to perform this integration is the most effective method for defining complex geographical relationships and for computing optimal routes.

Through GIS, a limitless number of variations related to the original route can be evaluated, and the most effective chosen. Within a GIS, a company can overlay relevant criteria to define the most effective route or prospective alternatives, and perform *what–if?* scenario and risk analysis. Once a decision has been made, GIS is also a perfect tool for communicating the decision and the reasons behind it through an easy-to-read map. While GIS provides data in a consumable medium, the technology also provides solid technical evidence that a planned project will meet the necessary compliance criteria.

Operations and Maintenance

The objective of the operations and maintenance team is to keep the pipeline running smoothly. To do this effectively, these teams manage both the physical assets and information about the assets that is used to make operating decisions. As the information is spatial, GIS systems are used to create, remove, update, and manage digital records about the physical assets. By integrating a GIS-based mobile solution into the operations workflow, information that was once collected on paper forms can be captured digitally in the field and used quickly to update the centralized GIS database. A mobile workforce can capture information associated with asset mapping, inventories, condition assessment, maintenance activities, and inspections.

Pipeline alignment sheets have long been the mapping medium for pipeline teams. GIS has revolutionized their generation by turning a largely manual procedure into an automated process. By creating a map template and assigning digital data such as topography, material, survey notes, or pipeline inspection gage (PIG) inspections to sections of the sheet called bands, operators can use GIS to quickly create pipeline integrity maps, regional drawings, easement maps, maintenance plans, and dig sheets.

Integrity Management, Risk Assessment, and Regulatory Compliance

Three of the most critical uses of GIS are for integrity management, risk assessment, and regulatory compliance. Using GIS, organizations can integrate and align multiple types of inspection information with traditional pipeline asset and spatial data to answer vital questions about integrity, safety, efficiency, and operations. GIS can help reduce corporate risk, decrease operating costs, and improve overall asset optimization.

While individual direct assessment or inline inspection records can be reviewed for integrity management purposes without the use of a GIS, it is almost impossible to integrate successive inspections to assess chronological decay. Each individual record uses the measure associated with the inspection tool's odometer or created by a survey crew. To make these records correlate, a GIS can be used to align data from the individual inspection runs with the pipeline centerline. An operator can then use the GIS query functions to quickly identify regions with anomalies/corrosion that require further investigation, or to produce detailed alignment sheets for more detailed analysis by an integrity engineer.

Risk assessment focuses on risks stemming from physical or legal causes (natural disasters or fires, accidents, deaths, lawsuits) or any other interruptions in an organization's ability to operate. It is a way not only to increase public safety but also to optimize pipeline design, operations, and maintenance. The GIS modeling interface is used to create and process a risk algorithm based on internal and external threats and consequences. GIS is used to connect elements of the risk algorithm with the appropriate data sources and then display and analyze the results.

While pipeline regulations vary greatly between countries, they are all geographic questions. Thus, a GIS is the most effective tool to achieve and document compliance. GIS can be used to calculate high-consequence areas, define the pipeline operating class, assess the pipeline burst radius, indicate potentially affected populations, etc. In each of these cases, geographic data is used in conjunction with a spatial model to calculate the measures and information necessary to meet governmental or corporate requirements.

Land and Environmental Management

Pipeline operators have assets that stretch across large sections of land, and they often manage a larger number of land records than those of many cities or counties. Depending on the location, these lands may be owned or leased, and the pipeline operators have to manage the legal and spatial parcel/right-of-way information. Commonly an operator purchases or receives sets of parcel, ownership, and address information from regional authorities or data resellers. When stored in a GIS, land management data can be updated seamlessly, accessed quickly, and shared with all departments in the company. With land records easily accessible in a GIS, the data can be used in numerous activities such as route planning, regulatory compliance, and mapping activities to reduce operational risk.

In the event of a rupture or spill, GIS plays a vital role as the pipeline company works to assess the problem quickly and later report on the situation. In a liquid pipeline, using topographic variables and information about the viscosity of the product, a GIS can be used to define a spill area and determine which sections of a population or the local water bodies and animal habitat could be impacted. If gas is released, GIS can calculate the plume dimensions and extent based on climate and product information so that communities can we warned and evacuated. Regardless of the situation, with spatial modeling, data integration, and mapping/reporting capabilities, GIS enables an operator to easily and effectively complete the tasks of evaluation and response.

GIS for Electric Transmission

The electric transmission system delivers electric power from generating plants to distribution substations. It is a complex system of overhead wires, underground cables, towers, heat exchangers, power transformers, high-voltage circuit breakers, and extensive control and monitoring systems. It is geographically distributed. Therefore, GIS is a natural tool to help operators manage, maintain, and operate this critical infrastructure.

The planning process for a transmission system is a geographic exercise. Transmission planners must have extensive data about the land features that may be considered for a possible new transmission line. They must be able to analyze all the data about the land to be able to determine the best route for a new transmission line and the associated equipment, including perhaps extensive switching equipment. GIS provides the tools to perform the spatial analysis that takes into account a multitude of factors in transmission siting, such as:

- Land topography: elevation, vegetation, soils
- Environmental
- Legal
- Cost: GIS can determine the lowest cost path based on number of turns and distance
- Obstructions: rivers, highways, cities
- Maps of opposition areas
- Weather
- Hazards.

GIS used for transmission routing takes many spatial data sources, optimizes the choices, and presents the results in the form of map that clearly shows the results of the analysis.

After the transmission system is built, GIS helps transmission operators prioritize maintenance activities and upgrades based on different factors, such as salt contamination, fire hazards, vandalism, and lightning strike density. This means that everyone within the company can access all the relevant information within a single view. It should be possible with enterprise GIS for anyone working with transmission to visualize the system and collect the results of:

- Helicopter inspection surveys of transmission line equipment
- Infrared surveys showing hot spots
- Encroachment in the right-of-way, such as swimming pools, abandoned cars, and metal sheds
- Lightning strike assessments
- Salt or smog contamination results
- Grounding surveys
- Access points along the transmission line right-of-way.

In addition, anyone within the utility can view recent maintenance activity, earthquake activity, areas where trees grow fast or slow, areas of higher than normal rainfall, and areas of significant environmental or cultural sensitivity. Each factor can be viewed in terms of proximity to oil-filled transmission equipment.

The geographic view highlights common points of failure, such as where transmission lines cross highways, railways, and waterways, or where critical transmission lines are near each other. It is important to know the legal attributes of the land (easements, licenses, leases, and their expiration) that lie within the transmission right-of-way, as discussed later in the chapter.

28.4.6 GIS for Distribution

Gas Distribution

Geography is a factor in every aspect of a gas utility's business. GIS integrates business data with geography for better analysis and decision making. GIS can help answer questions such as:

- How can I predict the portion of my system that could be damaged by third-party construction?
- What is the most efficient way to route field crews?
- How can we prepare to meet the demands of a growing community?

Below are some of the areas where GIS is commonly applied in the gas utility industry.

Asset and Facility Management

Gas utilities are spatial businesses with an extensive number of assets located in all quadrants of their operating area. Just as with other pipeline operators, gas utilities must know the status and condition of their assets at any time. A typical gas distribution system consists of assets such as main pipes and service pipes, valves, regulators, joins, and fittings. A map is the most effective way of locating and maintaining information about these individual assets. These assets are combined in a special type of GIS database called a *geometric network*. What makes the geometric network unique is that it not only maintains detailed information about all of the system assets but it documents how the features are connected—commonly called *connectivity*. Through GIS, updates are made to the database to keep asset and connectivity records current.

With a complete system inventory, operators are able to make informed decisions about their operations and maintenance programs. Analysts can monitor asset conditions to assist in infrastructure lifecycle planning, equipment maintenance, and asset replacement. Field crews can use mobile GIS tools to capture inspection information and quickly update the centrally stored database with as-built construction data. Engineers use a GIS to monitor supply, gas flow, pressure, and cathodic protection systems. The GIS-based mapping environment allows engineers to diagnose problems, prevent outages, and ensure corrosion protection.

Risk and Regulatory Compliance

As with transmission and gathering pipelines, there is risk associated with operating a distribution system. Organizations are required to adhere to regulations framing how they operate. Distribution risk assessment requires an organization to balance the physical/operational threats and consequences of systems against maintenance and monitoring programs. This allows operators to define areas with high risk and to perform the proper activities that ensure public safety. While risk requirements and operating regulations differ greatly from one region of the world to another, companies use GIS analysis and modeling to meet them. GIS provides the intuitive and logical framework for creating models, connecting operational or spatial data, and visualizing and communicating results.

Engineering and Design

As more requirements are placed on gas utilities and the technology of the equipment advances, the design of systems becomes more complicated. To engineer a system properly requires better access to accurate, current, and precise information. Engineers are now creating GIS-based designs on top of datasets such as satellite imagery, land use, parcels, and planned/associated infrastructure. GIS optimizes the engineering and design process and improves cost estimation by exposing the existing geography and landbase features to a design engineer. GIS streamlines the design process by using a rule base to automate steps in the design process, perform engineering calculations automatically, and link directly to purchasing, accounting, and work order systems. GIS allows utilities to predict development areas. Using the core GIS data and development plans, engineers can predict growth patterns and design appropriately for future system requirements.

One Call or Call Before You Dig

Even though underground pipelines are identified with pipeline markers, a major cause of pipeline accidents is third-party activity. The number one cause of pipeline leaks is damage caused by earthmoving and construction equipment or tools owned by parties other than the pipeline company. *One call*, or *call before you dig*, is the process of submitting construction requests and locating buried pipelines geographically prior to construction or excavation. It begins with an external source notifying a one-call or call-before-you-dig agency in the work area.

If an operator maintains an accurate one-call system, they can quickly identify whether crews need to be dispatched to mark company infrastructure. The ticket location is displayed in the GIS. The operator reviews the spatial coordinates in conjunction with the GIS data such as pipeline data, topography, site plans, etc. If it is determined that a pipeline is in the work area, the ticket and the digital information are used to route a crew to the site. The ticket is then cleared by the operator, and its status is communicated to the one-call agency.

Logistics and Workforce Management

Gas utilities have large workforces performing maintenance, construction, and meter-reading activities. GIS connects these field employees to the most current information using mobile or handheld solutions. Staff can also directly modify GIS information or input field-based design or as-built information, removing unnecessary data input and update steps.

GIS can also be used to manage and route a mobile workforce. GIS-based routing allows an organization to reduce labor and fuel costs, make efficient use of specialized equipment or highly trained employees, and properly manage staff allocations. Logistical solutions through GIS enable operators to optimize routes and workforce schedules, track assets and workflows, dispatch or reroute crews, and manage a mobile workforce.

Land and Environmental Management

Gas utilities generally rely on the landbase information that is created by municipal and regional agencies, supplemented with development plans from local planning agencies. These base layers are connected with customer and service databases to create a GIS layer defining the parcel structure and the customer base. This is used in many different types of activities ranging from infrastructure planning and gas service marketing to outage/emergency response notification and the detection of illegal service connections.

GIS can help address environmental issues related to construction activities and system operations. During all stages (planning, construction, operations, and maintenance), GIS is commonly used as the modeling and reporting tool for environmental and impact assessment activities. GIS provides the analytical framework used to integrate spatial data and determine the physical impact of an activity or accident. With GIS, organizations create detailed maps and reports required for evaluation, response, and government submissions.

Leak Detection and Outage Management

Using GIS, you can identify, isolate, and map areas of concern during a leak or outage. With a map interface, organizations can select valves or structures that need to be closed—referred to as a *valve isolation trace*—in order to separate a leak area from the rest of the network to mitigate loss. Organizations can also use GIS to trace the network to specify parcels and associated customers downstream of a main break, create leak reports for use in notification and pilot relight activities, and reroute resources to ensure that the number of people impacted by an outage is as small as possible. A GIS database can also be used to catalog and analyze all system leaks to create a better diagnostic system for predicting leaks. The results of this analysis are used in risk assessment and integrity management planning.

Electric Distribution

GIS is used extensively in the electric distribution business, although mostly as a tool for network documentation. The term *automated mapping/facilities management* (AM/FM) was used to describe the early attempts at digital mapping in electric companies. The problem with AM/FM systems is that they were designed to automate the process of making operations maps. Utilities routinely employed map makers to detail the location of their distribution facilities on linen manually with India ink or later using regular pencils on plastic transparency material. This linen then later plastic transparency sheets were used to create the old-fashioned blueprints. The goal of the old AM/FM systems was to attempt to replicate these manual drawings as exactly as possible. Over time, those systems were used to capture the connectivity of the electric distribution networks. The data from those older digital mapping systems often served as a data source for outage management systems.

Rarely were those systems used for analysis. More rarely still were those systems integrated within the fabric of the mainstream information technology (IT) infrastructure. Many of those systems did not use established mapping conventions or common map formats. Many utilities today continue to use legacy AM/FM systems. A lack of a standardized coordinate system has inhibited advanced spatial analysis, precluding companies from the use of imagery or GPS devices. One utility reported that when they attempted to overlay their digital mapping image with a map produced via light detection and ranging (lidar), their mapping system was off by more than 60 m.

Many utilities have either recreated their system using standard mapping conventions or they have used a process called *conflation* to adjust nonstandard mapping systems to standard map projections. Utilities are beginning to use GIS for analysis, performing tasks such as risk and vulnerability analysis, and using data from within and outside the utility.

Essentially, the electric distribution GIS consists of the following layers:

- Landbase or base map (or cadastral): This layer shows land elements such as streets, land features, property lines, water features, and other points of interest. It could include a variety of imagery such as three-dimensional (3-D) building features. From an electrical perspective, this layer is static.
- Structures: This layer includes common non-current-carrying structures such as poles and manholes.
- Current-carrying devices: This layer includes cables, wires, transformers, switches, and monitoring devices.
- Customer: This layer details the locations of customer meters.

Today, GIS is commonly used for nearly all facets of the workflow described earlier in this chapter. What is less common, but gaining in popularity, is the use of enterprise GIS as a platform for many distribution workflows. Much insight is gained by analyzing storm assessment, logistics for crew dispatch, tax determination, street light management, and billing.

28.4.7 GIS for Retail

Gas Station Business Analytics

The location of a gas station or commercial fuel outlet is based on geographic analysis. Operators work to balance economic and public factors to define the optimal location. Retailers assess the direct construction cost, proximity to population centers, distance from their own stations, proximity to competitors, regional income, traffic statistics, easements of existing utilities, the magnitudes of environmental pollution parameters, etc. While these are the company's internal concerns, they are also asked to minimize the impact that construction and long-term retail will have on traffic flow, accident instigation, and air/noise pollution.

Each consideration is a geographic measure or a spatial feature. Although retail analysis can be completed without the use of GIS, many of the datasets are only available as spatial data. Many retail operators have developed GIS-based models that combine these factors and consequences to define the optimal location. The analysis also generates the materials and maps required for construction and environmental approval.

28.4.8 GIS for Energy Consumers

Marketing

Utilities commonly encourage businesses to locate within their service territory for several reasons. They want to increase revenue of course, but they also want to make sure their facilities are fully utilized. If a gas company invested in a new pipe distribution system, it would want to make sure that the pipe is delivering gas close to its capacity. Utilities will encourage development in areas where they have excess capacity and discourage development in areas that would require significant capital expenditures to serve new load. The utility has an analytical problem: find a suitable site for a potential customer that maximizes the use of facilities but minimizes infrastructure expansion costs. GIS provides the tools to solve this problem.

For a given potential industrial customer looking for a site, GIS seeks land or an applicable vacant building that is the right size, is zoned correctly, is available for sale or lease, and would be served by an existing utility facility capable of meeting the proposed customer's energy demand. Additional considerations might be access to transportation facilities, attractive amenities, developed roadway infrastructure, and any other spatial considerations that the potential customer might require.

GIS quickly assesses various options and presents the results in the form of a map, perhaps grading each proposed site on the basis of a weighted assessment.

Smart Metering

One of the key elements of any smart grid initiative is the implementation of smart meters that identify consumption in detail so that consumers can monitor their energy use in real time. Smart meters in conjunction with infrastructure to monitor, control, and manage data is referred to as advanced metering infrastructure (AMI). The purpose of AMI is to read consumption data centrally, in contrast to the old practice whereby meter readers visited every meter. In more recent years, utilities have implemented meter-reading systems that collect meter information wirelessly. Meter readers can drive down streets and collect data at a short distance from a meter. Manual, automatic, and drive-by reading of meters still result in the meter being read once per meter cycle. Some utilities still read meters less frequently. Generally, older centrally managed systems only have the capability to read meters once per month. For the majority of customers, the only parameter read is the consumption of kWh or electric energy used for that month. Other than some industrial and larger commercial customers, utilities do not capture the actual demand at any given time, so utilities have no knowledge of when customers are using electricity.

AMI is different. Smart meters constantly monitor (or have the capability to monitor) consumption at any time. Thus, with AMI, utilities will understand customer use patterns throughout the day to better manage demand. Smart meters can also communicate with customer equipment and power down during peak load hours. Utilities could craft rate structures whereby customers would receive credits for participating in load management. Load balancing saves utilities from adding new power generation or distribution equipment, thus reducing overall greenhouse gas emissions.

During power failures, utilities will get immediate notification of a customer outage. This information can be used to help utilities automatically switch circuits to minimize the time and extent of outages. This process is called self-healing.

GIS plays a strong role in both the management and rollout of smart meters. By integrating data from smart meters, utilities can see geospatial trends in customer consumption, make better decisions on equipment load, and query customer meter data to display the results spatially. In addition, GIS is valuable in the conversion of conventional meters to smart meters. Since smart meters need extensive communication systems, GIS can provide data management of the telecommunications facilities, help in optimizing the locations of the various sensors and switching devices, and provide the logistics support for implementation crews.

Customer Care

We often think of GIS as managing and modeling technical systems, such as gas mains, electric facilities, and oil wells. GIS is also a strong business tool. It helps businesses locate stores, find buying patterns, and route sales staff. Energy companies also have many of the same customer care and business needs. Knowing the demographics and behavior of the customer base can optimize business processes such as collections, the marketing of a new conservation product, or customer satisfaction tracking. Since customers are geographically distributed, GIS provides a way to manage customer data, analyze customer behavior, interact with customers, and assess results. The GIS platform provides an easy way to consume social media data. This, along with demographic data, gives energy companies new ways to better understand their customers, vendors, and regulators.

28.4.9 The GIS Platform and the Problem of Abstraction

Energy companies have used a number of information and operational applications for years. For example, electric utilities routinely simulate the behavior of the electric grid with:

- Balanced and unbalanced load flow studies
- Three-phase and single-phase short circuit analysis
- Dynamic, transient, and steady-state stability (both single phase and three phase)
- Insulation coordination
- Immediate, short term and long-term load forecasting
- State estimation
- Relay protection and fuse coordination studies
- Harmonic studies.

Each of these simulation studies requires four processes:

1. Data management
2. State management (different cases/contingencies that happen at different times)
3. Calculations
4. Presentation of the results.

The most complicated from a mathematical perspective is the calculation. For example, a balanced load flow calculation requires the solution of a simultaneous set of nonlinear differential equations. Yet, in all of the above cases, the algorithms for these solutions are well known. Digital solutions for these simulations have been around for decades. The problem lies with the other three processes. The difficulty originates from abstraction, habit, existing work flows, and organizational silos.

These simulations require that the network information be abstracted into a schematic-like (nodes and lines) representation of the network, but not exactly the same schematic. To overcome this problem, utility analysts continue to maintain a unique set of data for each application. So, planners performing load analysis maintain their private set of abstracted data while system protection engineers maintain their network dataset. To add more complexity, state management requires not just one set of network data that employees must maintain but many different versions of the network data. They create the network dataset for the as-is condition, the various contingency datasets, and, of course, a whole range of datasets for future states (and their contingencies). So, for each of the above applications, there are likely to be hundreds of network datasets.

The problem then happens when something new happens. A transmission line project is canceled or delayed. This means that all of the datasets have to be corrected to account for just one change. Since the datasets are typically maintained by separate groups of people, the work to keep these datasets in sync is huge. Also, since the process of updating the datasets is mostly manual, the possibility for error increases.

Historic GIS and AM/FM systems abstract network information in a different way. Instead of abstracting the data schematically, they abstract the data into a two-dimensional map framework. While they maintain connectivity and location, most GIS simplify the elements into points, lines, and polygons. GIS does include the structural elements, such as poles and structures. Yet, in many GIS implementations, not all the structural elements are captured.

However, what the GIS does extremely well is data management, data presentation (visualization), and representing the various states of the network.

In addition to simulation systems and the GIS, utilities also use other IT systems such as enterprise asset management (EAM), work management systems (WMS), and reliability centered maintenance (RCM) to manage their network assets. These systems tend to abstract the network into data files organized around individual assets without regard to connectivity or location or complex electrical modeling.

Utilities also deploy hybrid GIS/work management/SCADA and network analysis systems, such as advanced distribution management systems (ADMS).

One common characteristic is the binding of data management, calculation, state management, and visualization into each of these systems and applications. What none of them perform is the actual abstraction itself. Instead, all of these systems (including older GIS) manage their data at the abstracted level.

The promise that the GIS platform holds is to manage the network data at a much higher level of abstraction. Since modern GIS platforms now embrace 3-D technology, the notion is that the highest level of abstraction is an accurate representation of the real physical network.

In this case, data management only needs to happen at this level in the GIS, not in each application. In addition, the state (the various projects and versions) can happen at this level as well. The abstraction can happen at the application level (not at the data management level). The workflow would look like this:

1. All network data maintained in one place in the GIS. This includes all the data needed for every application—not one set of data tailored to a particular application.
2. All states of the network managed (as they are now) in the GIS. Note that modern GIS have advanced means of managing projects. Note that there is only one database for all versioned data.
3. The GIS publishes web services based on the needs of a particular application.
4. The application normalizes and further abstracts the data into the appropriate form that matches the needs of the application.
5. Other systems, such as a meter data management system, a load forecasting system, or an AMI (advanced metering infrastructure system), would then publish its data via a web service to the applications (or the GIS). For example, a load flow system needs customer loading data.
6. The applications publish their results in a web service. The GIS can immediately reference the web service and take whatever action is needed.
7. The results of the applications can then be widely shared with anyone, on any device.

This puts the GIS at the center of data management, state management, and visualization. It then serves the needs of more complex applications. The same concept applies to gas utilities (and water utilities).

28.5 Summary

The GIS platform is a critical information technology platform for energy companies and utilities. GIS is also used by companies that serve the energy industry, such as engineers, designers, and contractors. Energy companies deal with assets, people, and processes through strong spatial dimensions. Utilities have assets and customers distributed nearly everywhere in the world. Gas companies have pipes up and down city streets. Oil rigs are spread throughout the world. The work of energy companies and utilities is always linked to geography.

GIS solves problems using four patterns of GIS: data management, analysis, mobility, and operational awareness. Each pattern is different but interrelated. From a GIS perspective, each pattern has a unique set of workflows and tools. GIS is involved in all the major energy infrastructure workflows.

Energy companies and utilities have distinct data categories and can use GIS to visualize layers of information on land, facilities, and structures. GIS organizes features into layers to better manage data and to facilitate the visualization of the data and the relationships of datasets to many others.

The energy business value chain consists of a number of discrete business units or master processes. These master processes are exploration, development, gathering, production, transportation, distribution, retail, and consumption. GIS plays a strong role in each.

GIS is pervasive in the energy business because geography plays such a strong role. Energy companies use GIS to save money, improve market share, help organize workflow, serve customers better, and collaborate better with their communities.

Further Reading

1. Hohl, P.: Delivering Water and Power: GIS for Utilities. Esri Press, Redlands (2021)
2. Meehan, B.: GIS for Enhanced Utility Performance. ArcTech House, Norwood (2013)
3. Meehan, B.: Modeling Electric Distribution with GIS. Esri Press, Redlands (2013)
4. Meehan, B.: Spatial Data Management – Changing Behavior. Energy Central, Aurora (2021)
5. Meehan, B.: Utility Network Pays off Your Technical Debt. Utility Dive, Washington (2021)
6. Hohl, P.: Electric Vehicles – All That Is Left to Do is Everything. Energy Central, Aurora (2021)
7. Meehan, B.: CIM Cimplified, Parts 1, 2 and 3. Energy Central, Aurora (2021)
8. Meehan, B.: Taking the Gut Feel Out of Customer Engagement. T&D World, (Endeavor Media), Nashville (2020)
9. Hohl, P.: Two Big Ideas in Customer Care. Energy Central, Aurora (2021)
10. Hohl, P.: Two Things to Know About Digital Twins. Energy Central, Aurora (2021)
11. Meehan, B.: Optimizing Mobile Workforces. Energy Central, Aurora (2021)
12. Meehan, B.: 1. Asset Management Starts with Great Data. Utility Dive, Washington (2020)
13. Meehan, B.: GIS Helps Workers Do Their Jobs. Energy Central, Aurora (2021)
14. Meehan, B.: The Perfect Storm for Utilities, Weather, Sparks and COVID-19. Energy Central , Aurora (2020)
15. Meehan, B.: It's a Cliché, TLM Is a Mess. Energy Central, Aurora (2019)
16. Meehan, B.: GIS Is the Foundation for Grid Modernization. Energy Central, Aurora (2018)
17. Meehan, B.: Why a Tough, Smart Grid Is Critical to Smart Communities. Energy Central, Aurora (2018)
18. Meehan, B.: Is Your Utility GIS a Digital Twin – Or a Digital Mutant? Energy Central, Aurora (2018)
19. Meehan, B.: Digital Regression vs Digital Transition vs Digital Transformation. Energy Central, Aurora (2018)
20. Meehan, B.: Heavier Than Normal Call Volume. Energy Central, Aurora (2018)
21. Meehan, B.: The Internet of Where. Energy Central, Aurora (2018)
22. Meehan, B.: Why 'The Science of Where' Matters for Utilities. Energy Central, Aurora (2017)
23. Clay, D.E., Shanahan, J.F. (eds.): Nutrient Management for Energy Efficiency. CRC, Boca Raton (2011)
24. Thomas, C., Warren, H., Meyertholen, E.: Mapping the Nation: GIS for Federal Progress and Accountability. Esri Press, Redlands (2011)
25. Ross, C.E.H., Sloan, L.E.: Terra Incognita, A Navigation Aid for Energy Leaders. Pennwell, Tulsa (2007)
26. Meehan, B.: Empowering Electric and Gas Utilitieswith GIS. ESRI Press, Redlands (2007)
27. Gaddy, D.E.: Introduction to GIS for the Petroleum I

William (Bill) Meehan William Meehan holds a BS in Electrical Engineering from Northeastern University and an MS in Engineering from Rensselaer Polytechnic Institute. He is a registered engineer and leads the global electric industry practice at Esri. Bill is also the author of the several books on GIS.

Robert Brook Robert Brook is an Industry Consultant who works with utilities to define and architect GIS-based solutions to critical business problems. Before his pursuing a career in consulting he filled key industry positions such as the Pipeline and Gas Utility Industry Manager at Esri and the Senior Director of Gas IT at PG&E.

Jessica Wyland Jessica Wyland is a journalist and writer. She currently researches and reports on how geographic information is used to respond to international energy concerns. In this capacity, Jessica works with utility companies, resource developers, and industry experts to promote and share advances in related technologies and practices.

Geographic Information Systems in Health and Human Services

29

William F. Davenhall and Christopher Kinabrew

Contents

W. F. Davenhall (✉)
Geomedicine Analytics Section, Davenhall Associates, LLC
Mount Dora, FL, USA
e-mail: billdavenhall@gmail.com

C. Kinabrew
National Network of Public Health Institutes
New Orleans, LA, USA
e-mail: ckinabrew@nnphi.org

Abstract

The chapter begins with a general overview of how GIS has evolved in the health and human services over the last several decades and provides readers with important definitions and descriptions (Sects. 29.1 and 29.2). Sections 29.3 and 29.4 uncover how GIS became an important tool for epidemiologists in the work of tracking infectious diseases and perfecting the study of population health. Readers will also learn that GIS adoption by hospital marketers and planners in the United States accelerated rapidly after 1970, when US Census data became relatively freely available in digital form. The importance of the legendary work of the Dartmouth Health Care Atlas Project and its founder Jack Wennberg is also introduced. In areas where high GIS adoption rates occurred, such as in public health, we feature key applications such as immunization management, disease tracking, outbreak analysis, disease surveillance, syndromic surveillance, emergency preparedness and response, community health assessment, environmental health, chronic disease prevention, and animal and veterinary health. Section 29.5 describes how GIS education has expanded across the academic fields of public health, healthcare administration, and social services. It is pointed out that the material presented in this chapter is not intended to be an exhaustive examination of the history of GIS but, rather, a brief introduction and overview that will generate further interest and self-discovery. Section 29.6.6 considers the role of GIS in response to the COVID-19 pandemic.

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_29

Keywords

geomedicine · human services · health services · healthcare · veterinary health · epidemic disease control · administration · public health · informatics · COVID-19 pandemic

The increasing utilization of geographic information systems (GIS) technology over the past several decades has transformed health and human services (HHS) and given *new eyes* to public health professionals, researchers, and hospital and health system employees, as well as the public they serve. While maps and spatial analysis have a long history in the field of health and human services, the early twenty-first century is a time when health professionals and the public have many powerful spatial analytic tools at their disposal. This chapter begins with definitions of health and public health (Sect. 29.1), followed by a brief history of geography and GIS in health and human services. It provides numerous examples of health and human services challenges and how GIS benefits them. The chapter includes a special focus on GIS for several specific disease programs. It concludes with a summary of GIS in HHS education. Overall, it takes a selective approach due to space limitations.

29.1 What Is Health?

Everyone agrees that health is important, but what is health? Shortly after World War II, the World Health Organization (WHO) defined health as *a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity* [1]. This holistic view incorporates total well-being under the concept of health. Over the years there has been debate over what should be *in* and what should be *out*, but in practice it has been a challenge to measure the current WHO definition of health due to its broad scope. While the world has an International Classification of Diseases (ICD) [2], it does not yet have an international classification of health. So while nations have agreed on how they will classify diseases and other health problems, and the ICD provides a basis for national health statistics that can be compared across nations [3], in the early twenty-first century there is not yet a worldwide system for measuring health.

No matter how governments and organizations may define health (positively or negatively, holistically or narrowly), improving health appears to be a universal priority. Several decades after the WHO formulated the above-referenced definition of health, the Lalonde Report (a working paper from Canada's Minister of Health and Welfare) made quite an impact on the global health community. It noted that most of society's efforts and expenditures to improve health had been focused on the health care organization. Yet, analysis revealed that the main causes of sickness and death in Canada were rooted in three other elements: human biology, environment, and lifestyle [4]. The paper urged society to reconsider spending vast sums of money *on treating diseases that could have been prevented in the first place* and introduced a framework of four broad determinants of health:

1. Human biology
2. Lifestyle
3. Environment
4. Health care organization.

The Lalonde Report recognized that the existing health care delivery system could do *little more than serve as a catchment for the victims* of the other three determinants. It propelled the field of health promotion and influenced many significant initiatives to prevent morbidity and mortality from the major determinants of health (highlighted later in this chapter).

29.1.1 What Is Public Health?

Defining public health is also a difficult task. *Winslow*, one of the leading figures in public health history, defined public health in 1920 as [5]:

> … the science and art of preventing disease, prolonging life and promoting health and efficiency through organized community effort for the sanitation of the environment, the control of communicable infections, the education of the individual in personal hygiene, the organization of medical and nursing services for the early diagnosis and preventive treatment of disease, and for the development of the social machinery to insure everyone a standard of living adequate for the maintenance of health, so organizing these benefits as to enable every citizen to realize his birthright of health and longevity.

Other definitions have been proposed over the years, but Winslow's has stood the test of time. Two key differentiating factors between public health and clinical medicine are:

1. Public health focuses on prevention (as opposed to *sick care*).
2. Public health takes a community or population-based approach, instead of focusing at the individual level.

The *organized community efforts* and *social machinery* in Winslow's definition include not only governmental public health agencies at the national and local levels, but also a broad array of nongovernmental organizations (NGOs), universities, and other partners.

29.2 A Brief History of Geography and GIS in HHS

Merriam-Webster defines geography as [6]:

> a science that deals with the description, distribution, and interaction of the diverse physical, biological, and cultural features of the earth's surface.

One of geography's subdisciplines is medical geography, which [7]:

> focuses on patterns of disease and death – of how diseases spread ... and how variations in morbidity and mortality rates reflect the local environment.

A recent editorial in the *Journal of the American Board of Family Medicine* [8] reminds us that as early as 400 BC, Hippocrates said that one's health depends on the air one breathes, the water one drinks, and the environment in which one lives [9]. In this sense, geography has always been a part of individual health and public health. Maps and mapmaking have helped the medical and public health community understand the role of geography in outbreaks and pandemics of plague, cholera, typhoid, malaria, COVID-19, and numerous other infectious diseases. The study and control of epidemic spread at a variety of appropriate scales is inherently geographical [10]. Maps and mapmaking have also helped health officials and researchers analyze and understand chronic diseases such as cancer and heart disease, which represent a growing disease burden in developed nations as well as low-income countries. Whether one is looking at electronic web-based maps and globes portraying the COVID-19 pandemic ... or a 1694 map of the plague ... one thing is clear. Depending on the mapping and spatial analysis capabilities of the time and place, public health and medical professionals have utilized geography in their work. More recently, geography is becoming embedded in routine business processes of health and human services organizations (e.g., route planning for a fleet of vehicles).

In *Cartographies of Disease, Koch* describes many interesting chapters in history of medical geography, both before and after *Finke*'s *Medical Geography* [11, 12]. The formal discipline of medical geography started as early as 1792 ... Koch suggests that mapping and mapmaking *let us study the interpretations of selected aspects of spatially grounded, interrelated processes* [11, p. 6]. As we think about classes of discrete events (e.g., people with specific illnesses or symptoms) in the context of potentially relevant data (e.g., location, income, exposures) we are better informed to demonstrate or disprove a causal relation [11, p. 5].

Most of the works *Koch* describes in [11] are not simply disease maps but, rather, *mapped arguments about disease incidence and the environments that produced it* [11, p. 31]. The maps listed in Fig. 29.1 range on a continuum from descriptive to highly analytic. Some maps were trying to analyze the root causes of a local outbreak of disease at the scale of neighborhood or local jurisdiction, other maps were charting the progression of a disease around the world. Some maps were showing risk for a particular disease based on known causes (e.g., malaria and mosquitoes), others were showing the distribution of a particular disease at a national level (e.g., cancer). It is very clear from Koch's work that the history of medical mapping is tied not only to advancements in medical science and public health but also to advancements in the technology of mapping.

Much attention in the history of medical geography has been focused on John Snow, the *father of epidemiology*. Snow mapped the relative density of cholera deaths in relation to water sources, as part of a process identifying the infamous Broad Street pump (Fig. 29.2) as a likely source of contaminated water [13]. Koch reminds us that this work by Snow was part of a larger movement for better statistical and graphic analysis, which culminated in a series of extraordinary studies of cholera outbreaks in the 1850s, including the work of Snow (p. 38). While Koch focuses significant attention on John Snow, he also describes many other chapters in the history of medical geography, before and after the 1850s, and beyond London. Reading this history, it becomes clear that over the last few centuries mapping has helped physicians, sanitarians, boards of health, and high-level health officials (such as the Surgeon General of the United States) contain disease outbreaks and understand the worldwide spread of a disease. However, perhaps more provokingly, Koch makes the case that *maps and spatial analysis were critical in establishing the notion that public health was in the public health interest*, hence the rise of public health and sanitary science as a discipline. One notable event in this history is the International Sanitary Conference of 1874 held in Vienna. During that conference, maps of the spread of cholera were shared among attendees from many nations, well before the advent of the today's International Health Regulations and the regional and international Health GIS conferences that take place routinely each year.

29

29.2.1 The Early Years of GIS in Health and Human Services

In 1970, *Cline* commented that aerial photography was already in use by some epidemiologists for demographic purposes, the selection of random or systematic population samples, and to describe study areas [15]. He rightfully predicted that there would be future uses of remote sensing techniques within epidemiology. Aerial photography had been used in the past to identify and map disease vector habitats (e.g., as early as 1949 when *Audy* published a study on the distribution of scrub typhus in parts of Southeast Asia [16]). Much of this work was manual in nature, and the advent of GIS technology over the intervening decades facil-

Fig. 29.1 Timeline of medical geography: Highlights from *Cartographies of Disease* (adapted from [11])

1700 — 1800 — 1900 — 2000

- 1694 map of plague in Bari, Italy
- 1792 Finke's *Medical Geography*
- 1798 map of yellow fever in New York City
- 1819 map of yellow fever in New York City
- 1821 map of typhoid fever in New York City
- 1820s mapmakers add demographic data to map surface in an attempt to determine the nature of disease outbreak
- 1832 map of progress of cholera around the world
- 1832 map of cholera in Exeter
- 1836 map of cholera in Hamburg, Germany
- 1840 map of hernias in France (good example of medicostatistical mapping)
- 1841 map of sanitation in Dublin, Ireland
- 1850 map of cholera in London
- 1854 map of cholera in South London
- 1854–55 numerous maps of cholera in Soho
- 1856 world map of health and disease
- 1875 maps of cholera in New Orleans, Memphis, Nashville and USA
- 1876 map of cholera in East Africa
- 1878 map of offensive odors in Boston, MA
- 1885 maps of typhoid in Plymouth, PA
- 1885 map of sever outlets in Harrisburg, PA
- 1886 map showing distribution of population in State of Pennsylvania (in first annual state report on health)
- 1892 map of smallpox epidemic in Western Australia
- 1892 map of the distribution of cancer in British counties
- 1901 maps from Sedgwick's *Principles of Sanitary Science and the Public Health*
- 1925 map of sandfly fever in Peshawar, India
- 1950s maps from the Welt Seuchen Atlas
- 1960 map of cancer in Horrabridge, England
- 1961 maps from May's *Studies in Disease Ecology* (including world distribution of important malaria vectors, and a worldwide schematic of plaque foci)
- 1960s maps of the death rate in England
- 1962 map of Burkitt lymphoma in Africa
- 1969 map of schistosomiasis in South America
- 1971 maps of influenza in England
- 1972 map of bronchitis in Leeds
- 1977 map of hepatitis in Tasmania
- 1986 map of influenza in the United States
- 1988 map of malaria in India
- 1988 map of leukemia in India
- 1980s–90s maps of outbreaks of AIDS in Ohio and US
- 1989 map of HIV in Eastern Tennessee
- 2000 CDC releases EpiMap to display data from EpiInfo as maps
- 2002 map of AIDS in Thailand
- 2002 map of dengue fever in Pennathur, India
- 2004 map of AIDS in the United States

itated many epidemiological investigations that would have not otherwise been possible.

The launch of ERTS-1 (Earth Resources Technology Satellite) in 1972 provided the first opportunity to acquire global remote sensing data on a regular basis [17]. Through the 1970s and 1980s there was an increasing number of research projects utilizing remote sensing and GIS to investigate the epidemiology of diseases. In 1990, Louisiana State University (LSU) hosted the first International Conference on Applications of Remote Sensing to Epidemiology and Parasitology. LSU has become a collaborating center of the WHO in this field.

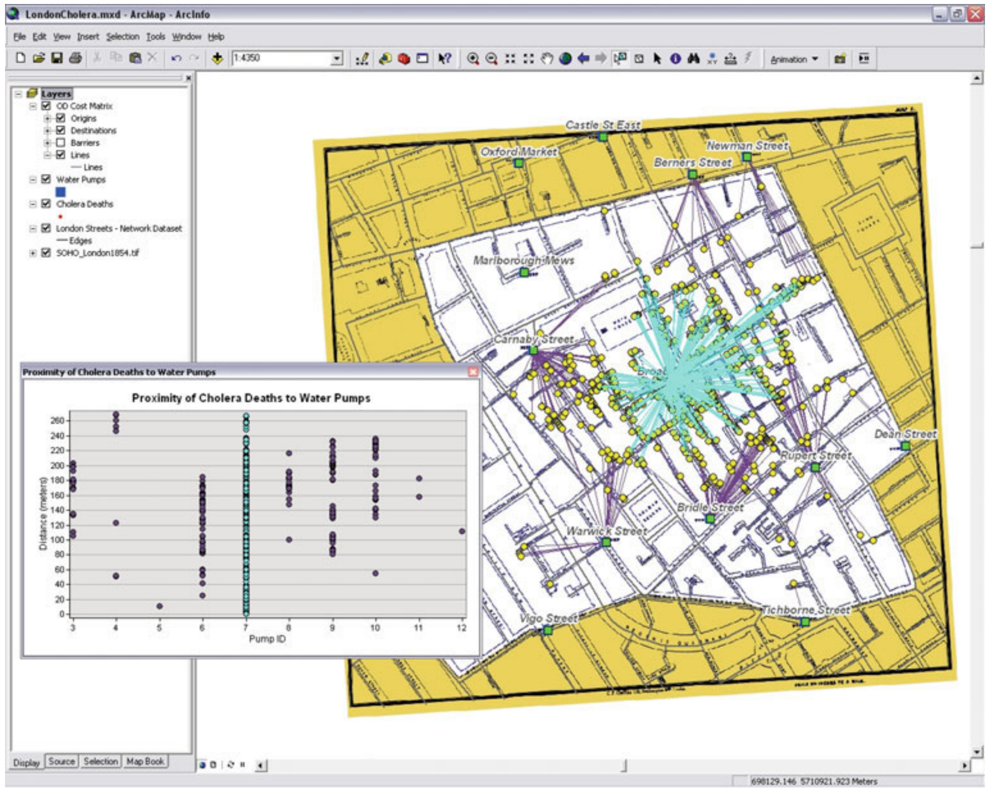


Fig. 29.2 John Snow mapped cholera deaths in proximity to the water pumps in London's Soho district in 1854. His map identified the likely source of the cholera outbreak as the Broad Street pump. Upon the pump handle's removal, cholera cases tapered off substantially (adapted from [14])

One of the first uses of GIS in hospitals occurred shortly after the release of the 1970 Census of Population in the United States. The US Census raw data was not available to the general public before 1970. However, in 1970, the US Census Bureau granted public use access to its raw data collection. This meant that for the first time researchers and commercial entities could work with relational data for small geographic areas, such as ZIP Codes, census tracts, and block groups. By liberating the finely grained census data health and human services organizations were able to begin to use the population data at the local levels.

Around the same time, the US Census Bureau released two important products that would forever change the way GIS or computer mapping programs would be focused. The programs were called Admatch and DIME. Admatch was a computerized address management program that would link a street address to a specific geography; DIME was a master reference file called the Dual Independent Map Encoding. The combination of these two programs ushered in the first *industrial* strength geocoding technology (as we call it today), which is one of the central components of a modern-day GIS. However, one must remember that approximately 50 years later (i.e., in 2020) similar address level geocoding is still not available in many countries of the world.

In early 1973, the Ford Foundation Fund granted a non-profit organization called DUALabs (a data-use access laboratory) the resources necessary to understand how the Census Bureau's raw data tapes could be disseminated in a useful and efficient way. The DUALabs legacy was that it provided a business model for most of the commercial demographic companies in the USA and paved the way for mapping census data. Clearly, without the release of the public use files of the US Census of 1970, the commercial mapping industry of the USA would have been significantly smaller and would have taken much longer to become an established way to analyze geographic and demographic data.

In 1973, *Davenhall* et al. [18], using the software and data products available from the US Census Bureau, created the first patient distribution map of emergency room patients from St. Joseph's Hospital in Lexington, Kentucky. Creating the map required the use of the DIME file, the Admatch program, and the master patient history file from the hospital processed by an IBM 360 mainframe computer. Fortran was used by a programmer for a 6-month period to produce what was known as a dot density map in black and white depicting the dossier of patients in various small areas of the Lexington metropolitan area. The current value of the project in today's dollars would be about US $250 000. The one map that was created from taping about 30 8 1/2″ by 11″ sheets of paper was used by the hospital management to make an important decision in that community concerning expansion of the emergency room. In the course of the decade of the 1970s, great strides were made in the use of demographic and geographic data across a wide range of private and public interests, continuing to improve every aspect of GIS data and software use in the health and human services sector.

In the second half of the 1970s, the US Department of Health and Human Services (DHHS) sponsored a major demonstration and research project in Jefferson County, Kentucky, called the Human Service Coordination Alliance (HSCA). The HSCA project had as its goal service integration on a massive community scale – across several geographically contiguous counties in Kentucky serving the Louisville metropolitan statistical area (MSA) – a population of just over 1 000 000 people. The project consumed large amounts of agency service data, including beneficiary demographics, service delivery characteristics, and service performance measures. GIS (called computer mapping) was introduced into the project to illustrate the various market segments being served, as well as to provide visual communication of patterns and trends to health services agency leadership. The project included every concurrent human service activity from nutrition service to family service to personal addiction therapy.

More Affordable Software and Data

Through the 1980s, there was an explosion of relatively moderately priced software and data packages and services created by both public and private organizations to exploit the extensive data offering of the US Census. These new tools and data services greatly expanded the assessment (i.e., sensory) capacities of epidemiologists, as well as public health planners, environmental health professionals, monitoring and evaluation specialists, hospital marketing departments, and many others.

Graduating from Innovative to Standard Practice

In 1990, the Agency for Toxic Substances and Disease Registry (ATSDR), a sister agency of the Centers for Disease Control and Prevention (CDC) began using GIS to evaluate the demographic characteristics of populations around sites, identify features in the community that were relevant to the protection of populations, monitor health-related data and health complaints by community members, and overlay contaminant distribution with census data to characterize subpopulations that might be potentially susceptible to site contaminants due to their age, sex, race, or socioeconomic condition [19]. CDC's use of GIS expanded over time beyond ATSDR to the National Center for Health Statistics (NCHS) and many other centers, institutes, and offices within the agency.

Throughout the 1990s, health and human services organizations were able to acquire low-cost software and data bundles, which provided an attractive analytical solution. Public health agencies, hospitals, and social service agencies began to purchase these bundles and incorporate geographic thinking into the operational activities, largely at the desktop computer level.

The anthrax attacks and events of September 11, 2001 in the United States accelerated the CDC's and HRSA's nascent public health emergency preparedness programs as well as these agencies' cooperative agreements with state and local health departments across the USA. Given increased resources and responsibilities, health departments hired more epidemiologists and deployed robust information technology to support public health preparedness and response. Many health departments recognized the capabilities of desktop GIS software for public health preparedness. This was also the time frame during which both robust internet mapping software and mobile GIS software (e.g., Esri's ArcPad) were launched. Well before the advent of applications such as GoogleEarth, federal health agencies such as CDC and HRSA, as well as numerous state and local health departments, began publishing internal and public web-based data query systems (i.e., interactive health atlases) and web-based services locators. These agencies also enhanced electronic disease surveillance systems, immunization registries, and cancer registries with mapping and spatial analysis capabilities. However, it would take another decade for web-based situational awareness systems to emerge as standard practice.

29.2.2 The Early Adopters

As indicated in the above-referenced paragraphs, there were a number of different early adopters of GIS in health and human services, including epidemiologists, biostatisticians, vector control professionals, environmental health professionals, hospital planners, and human services professionals, among others.

Epidemiologists, the disease detectives of public health, saw the potential for GIS software very early since they consider the triad of person, place, and time in their investigations and analyses. Epidemiologists are responsible for responding to disease outbreaks, as well as other community health problems. Early on, they used GIS to collect data, geocode it, and then analyze the data, often as the result of a disease outbreak or a poorly understood community health problem. Having all this data in a GIS helped them understand disease diffusion, as well as to provide situational awareness to leadership (often through static maps), so decisions could be made regarding where resources needed to be deployed. GIS also provided analytical tools for exploratory analysis of potential sources of outbreaks. Static maps produced through GIS software showed study areas for epidemiological research papers and presentations. For most of these purposes, GIS was viewed as an analytical tool that served a narrow slice of the overall public health functions.

Biostatisticians working on large datasets (e.g., cancer registry) also appreciated the capabilities of GIS for data management, analytical functions, and visual display/reporting. Biostatistics is a branch of applied statistics and is concerned with developing and using techniques to summarize and analyze medical and biological data [20]. Biostatisticians are responsible for analyzing data and designing research studies. Early on, they used GIS to geocode data and then analyze large datasets, proactively looking for clusters or other spatial (aberrations) and also responding to calls for investigations. Biostatisticians have also used static maps produced through GIS software to share visualizations of their work with colleagues and the public.

Environmental health professionals, responsible for inspections and data collection covering large geographies – such as those working in vector control – saw the applicability of not only desktop GIS, but also mobile GIS. Such programs (sometimes independent of public health agencies) have workflows that are inherently geographical. Vector control staff were among the first to use mobile GIS on a routine basis. Environmental health professionals use desktop and mobile GIS to manage large inventories of facilities and sites under regulation ... and also to meet numerous regulations (i.e., an assurance function).

Individuals working in the marketing and planning departments of hospitals also recognized how GIS could help them analyze service/catchment areas and locate best sites for new services. In social services, information technology staff had the vision that helping case managers see the client in the context of their environment would help them actualize the core functions of social work.

The vignettes mentioned before highlight some of the pioneers, but today the story is vastly different. Most health departments in the USA and a growing number around the world, use GIS in their daily work, and it is not just epidemiologists searching for clues to difficult outbreaks or biostatisticians working on disease registries. Public health professionals use GIS to analyze chronic disease trends (e.g., heart disease, diabetes, cancer), analyze access to public health services (e.g., vaccinations), analyze the built environment, respond to natural and manmade disasters, and design community health communications programs. While there is still a heavy desktop GIS presence in health departments, increasingly it is also found residing within an organization's IT department, supporting web-based applications (both internal and public-facing). The most recent GIS servers allow GIS functionality to be deployed on mobile devices across the entire organization, so it is expected that not only vector control but also many other field-based programs will move in this direction.

29.2.3 GIS Starting in Hospitals

In hospitals, GIS gained acceptance in the early 1990s as an essential analytical tool for strategic planning and marketing, largely in the USA. Analytical studies of patient origins and resident destinations, health facility site locating, and market demographic analysis headed up the list of most useful applications. Also, many hospitals hired marketing and planning consultants who could bring the mapping tools and demographic tools with them. As hospitals came under greater pressure to slow the cost increase largely through reduction in work force in *nonessential* departments, computer mapping and demographic analysis were outsourced to large consulting firms. The deployment was almost exclusively desktop, and seldom was GIS part of an enterprise approach to IT.

It is worth noting that the use of GIS by public health and hospitals has been inextricably related. The hospital is still a large data generator for public health, while public health is a hospital's single most consumer of its complete data collection. Without each other's contributions of data and analysis, the adoption of modern GIS would have been seriously compromised. While hospitals in the USA are often not that close to the philosophical underpinning of public health, outside the USA this artificial separation is almost nonexistent. In the decades ahead, a greater synergistic relationship between those in acute care and those in prevention will drive greater utilization of GIS as resource allocation, community accessibility, and government accountability increase. There are many good examples of this data and analytic synergy, as highlighted in the following scenario.

An emergency room physician is exasperated by the number of motorcycle and traffic-related injuries coming through the hospital's emergency room. This physician, in collaboration with injury prevention staff at the health department, obtains traffic accident reports from the department of transportation, geocodes all the locations of incidents, and conducts several analyses (e.g., density analysis). Based on the results, the physician and public health department agree on several possible interventions, such as traffic calming measures, other road safety improvements, and an education campaign. In this scenario, GIS is a tool for the analysis, and the maps that are the outputs of the analysis serve as an advocacy support tool.

Perhaps the best example of the data and analytic synergy that GIS provides between hospitals and public health can be seen in the seminal work of the Dartmouth Clinical Evaluation Research Center, in conjunction with the Dartmouth Medical School (covered in detail in the following section).

29.3 Geography Is Destiny in Health

According to Goodman, *geography is destiny in health*. This quote is based on years of experience analyzing data from the Dartmouth Atlas of Health Care. In 1973, *Wennberg* and colleagues began studying small area variations in health care delivery [21]. Since then, Dartmouth researchers have been studying the differential utilization, cost, quality, and outcome of health care delivered to Medicare beneficiaries across the USA. Focusing on small area analysis, Dartmouth researchers began to use census data at very small geographic levels to examine the patterns in medical claims data, principally from hospital utilization. This legendary research began with a geographic promise that, somehow, the variations in how health care was delivered, priced, and evaluated varied by where it was taking place. According to *Goodman* [22]:

> unwarranted variation in health care is variation that cannot be explained by patient illness, dictates of evidence-based medicine, or patient preference

and

> unwarranted variation is caused by differences in the effectiveness and efficiency of health care delivery systems . . .

This body of research has subsequently come to help explain the wide variability in geographic terms, so that local interventions can be encouraged and government policy can be used to alter the most undesirable effects of the health system [23]. The study of small area variations in health care has become a priority in Europe and is one of the components of the ECHO project (Extension for Community Healthcare Outcomes) [24].

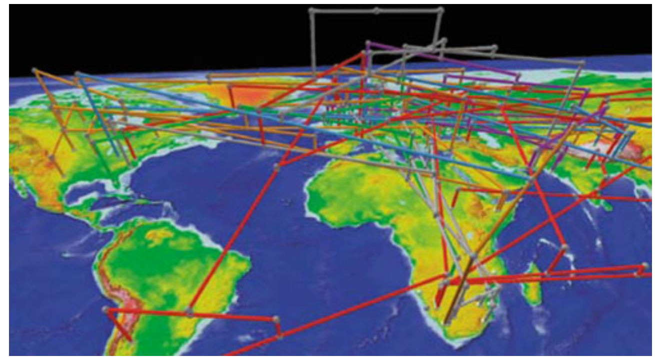


Fig. 29.3 Global distribution of *Bacillus anthracis* (the bacteria that causes anthrax). Matching anthrax strains are geographically connected (adapted from [27])

29.3.1 Growth of GIS in Health and Human Services

There are many drivers for the growth of GIS in health and human services, as detailed earlier in this chapter. Given the recent increased policy interest in the social determinants of health, it is worth discussing how GIS aids in the analysis of the four determinants of health referenced by *Lalonde* [4] several decades ago. Table 29.1 describes this notion at a high level. It is not meant to be a comprehensive catalog but just to make the point that the major geographical features of the Earth's surface (physical, biological, and cultural) mesh well with the determinants of health (human biology, lifestyle, environment, and health care organization).

Over the past 25 years, many leading health organizations, including WHO, World Bank, US Agency for International Development (USAID), and CDC have acknowledged the

Table 29.1 Determinants of health and their relation to GIS

Geographical feature	Determinant of health	Applicability of GIS
Biology	Human biology	Mapping genetic variations (Fig. 29.3) Mapping the human body [25, 26]
	Animal/plant biology	Remote sensing Controlling vector-borne disease Mapping food availability
Cultural	Lifestyle	Behavioral risk factor mapping Targeting health promotion Segmentation analysis *Place* studies
Physical	Environment	Investigating toxic exposures Controlling vector-borne disease Analyzing the built environment Analyzing the food environment
	Health care organization	Analyzing accessibility of health care services Routing vehicles Workforce studies Hospital-acquired infections Bed management Laboratory services

utility of mapping and spatial analysis in tackling some of the world's greatest health problems, including the COVID-19 pandemic, HIV/AIDS pandemic, malaria, tuberculosis, maternal and child mortality, and other devastating diseases. The majority of ministries of health (MOH) around the world, and all 50 state health departments in the USA have leveraged geographic information systems to assess and protect the health of the populations they serve. According to *Miranda* et al. [28]:

> Many GIS-based projects have been successful in supporting public and environmental health practice, including those investigating toxic exposure, vector-borne disease, health information access, and the built environment.

WHO states that GIS [29]

- Geospatial technology in the form of Geographic Information Systems (GIS) enables spatial representation of data to support better public health planning and decision-making.
- The global health and medical applications of GIS are numerous: Finding disease clusters and their possible causes; Improving deployment for emergency services; Determining if an area is being served adequately by health services.
- Geospatial data and techniques are also an effective tool to monitor progress and provide a strong basis for policy making to achieve the sustainable development goals and deliver the GPW 13 Triple Billion targets.

29.4 GIS Relevance to Public Health

As referenced throughout this chapter, GIS applications have been used extensively in health-related research and practice. This section takes a broader look at many of the major programmatic areas of public health, hospitals, and human service organizations and includes selected examples of the use of GIS to support.

- Immunization
- Disease surveillance, outbreak investigation, and syndromic surveillance
- Public health preparedness and emergency response
- Community health assessment, planning, and profiling
- Environmental health
- Chronic disease prevention and control
- Infectious disease
- Animal health/veterinary health
- Human services case management
- Hospitals and health systems.

The following section provides some additional context and examples of the application of GIS for disease-specific programs in public health.

29.4.1 Immunization

GIS has a long history of providing support in immunization campaigns around the world. Public health officials need to answer questions such as

- What are the vaccination rates in this area?
- Where should we concentrate our future efforts (i.e., are there pockets of need)?
- Where are the vaccines available?
- How can we best communicate information with partners and the public?
- Is our vaccination delivery network sufficient?
- Is there any clustering of adverse events?

Public health officials also need to communicate much of this information with partners and the public. Location information is critical to vaccination needs assessments, intervention planning, and visualizing/monitoring results. Following the COVID-19 and H1N1 pandemic, there has been an increased recognition of the utility of GIS for mass immunizations, better management of vaccine logistics, and the analysis and visualization of adverse events reporting.

Leading public health organizations around the world have validated the utility of GIS for immunization programs. Even in the most resource-constrained settings, WHO and UNICEF have encouraged hand-drawn mapping as a way for health professionals in local facilities and district health officers to understand unmet needs and monitor progress [30].

Health agencies have also used maps to visualize adverse events in hospitals and health facilities. In the United States, California's Department of Public Health has published a Hospital Adverse Event Map showing adverse events that hospitals reported per state regulations to the California Department of Public Health [31]. WHO has suggested that MOH develop a GIS system that can capture the geographical and frequency distribution of adverse events, *especially when the occurrence can be sporadic*. It suggests that the findings of the GIS may highlight the magnitude of some adverse events at the national level. An example from a specific country plan is also available [32].

USAID has supported a number of GIS capacity building initiatives around the world that have enhanced vaccination efforts. Researchers in Nigeria recently determined GIS microplanning is more cost-effective than traditional microplanning and that the higher costs incurred by GIS microplanning are worth adopting [33]. In 2010, the USAID-funded DELIVER project reviewed Esri's network analyst for use in developing countries. Their positive review suggests that *all* countries may consider using such solutions to improve the delivery and logistics [34] of numerous public health programs, including immunization. GIS has also been helpful in analyses of immunization coverage in conflicts, including Afghanistan [35]. UNICEF and the Global Alliance for Vaccines and Information (GAVI) gathered stakeholders in 2016 to review GIS for planning, monitoring and evaluation of immunization programs. There was consensus during the meeting that the integration of geospatial data and GIS in immunization programs offer significant potential to accelerate progress towards global vaccination goals, by facilitating the quantification and location of the target population for immunization, and by assisting in identifying missed populations [36]. In the United States, CDC's immunization program has encouraged state health departments to utilize GIS alongside immunization registries in order to analyze *pockets of need* (i.e., areas of underimmunization). CDC has also used GIS internally to detect spatiotemporal clusters of adverse reactions to vaccinations [37].

Health departments have used GIS applications to track stockpile shipments (including vaccines) in real time [38]. GIS has also proved to be a useful analysis tool in the administration of mass vaccination in local jurisdictions [39]. According to one presentation, GIS helped a local health department

- Determine who was treated (geocode patient information).
- Determine who needs to be treated (spatial join with population information, calculate vaccination rates).
- Notify those who need to be treated (activate automated telephone reminders in low vaccination rate areas).

29.4.2 Disease Surveillance, Outbreak Investigation, and Syndromic Surveillance

GIS has a long history of providing support to traditional disease surveillance. Most electronic disease surveillance systems now include a mapping module, so that analysts and public health leaders can visualize disease outbreaks on the map. When an outbreak is detected and public health personnel go into the field for additional investigation, GIS strengthens local data collection, management, and analysis. From the beginning, GIS provides a baseline for monitoring and evaluating outbreak investigation activities. Mobile GIS allows field personnel to leverage GPS devices to navigate more efficiently and quickly to locations for data collection. This is critical when time is of the essence. Surveillance of case locations is maintained more effectively, so the geographic progression of the disease is continually monitored. High-transmission areas (e.g., gathering places) or areas with environmental conditions ideal for disease vectors (e.g., standing water) are more easily identified when field staff have maps, imagery, and descriptive metadata at their fingertips.

GIS also facilitates targeting of prevention and control measures based on priority locations. For example, recent research has shown the effectiveness of implementing integrated vector control within a defined distance buffer of known dengue case locations [40]. GIS has also proved to be a cost-effective technology for controlling animal outbreaks (e.g., avian influenza) [41]. In other outbreak situations, GIS is a valuable tool for designing economically feasible population-based public health investigations (e.g., generating a spatially random sample during rapid needs assessments) [42, 43].

Syndromic surveillance systems have been used for the early detection of outbreaks, to follow the size, spread, and tempo of outbreaks, to monitor disease trends, and to provide reassurance that an outbreak has not occurred [44]. GIS supports syndromic surveillance not only as a visualization aid, but also in detecting abnormalities based on spatial queries. Examples include the use of dead bird reporting to predict West Nile virus in humans, and analyzing chief complaint data from hospital information systems to determine spikes in gastrointestinal or respiratory complaints.

With accurate location information, GIS can provide spatial visualization of complex relationships between cases, contacts, and objects in the environment in both time and space. Spatial visualization helps identify disease sources and the best implementation of countermeasure and response strategies.

29.4.3 Public Health Emergency Preparedness and Response

GIS has been used for decades in response to natural disasters such as floods, earthquakes, wildfires, and hurricanes. Following events such as 9/11, the anthrax attacks, and hurricane Katrina, the debate ensued regarding exactly what public health emergency preparedness was and how it should be measured [45]. Regardless of one's definition of public health emergency preparedness, it is clear that preparedness and response depend on location-based information, such as the location of incidents, where responders are, and where emergency services and health facilities are located. GIS supports emergency preparedness through

- Needs assessments and planning
- Evacuation route planning
- Modeling chemical spills
- Targeting emergency notifications
- Determining sites for points of dispensing (PODS)
- Enhancing the utility of emergency operations center (EOC) software.

During a public health emergency there will be dozens, perhaps hundreds, of response efforts going on simultaneously through both the public and the private sectors. Besides helping all responders and those directing them attain situational awareness (by seeing response activities *on the map*), flexibility is one of the most useful features of a GIS. By altering the planning assumptions that are entered into the GIS, public health officials can conduct analyses quickly and efficiently on any issue for which data are available [46].

> Using GIS (including data analysis pre and post mapping) helps to sensibly choose the best places to locate our limited resources in order to improve service to the public that we serve [47].

For disaster planning [48]:

> The analysis revealed many opportunities for improvement and as a result, GIS has now become a pivotal tool in the County's planning process.

Many public health authorities now have EOCs in order to increase situational awareness, collaborate with other first responders, and mount a more effective response. Increasingly, public health is seen as a first responder like fire, police, and public safety. There are many roles for GIS in health EOCs. The WHO Regional Office for the Western Pacific states that *information is the lifeblood of an EOC* in [49]. The guide suggests that an action plan should include supporting materials such as a map of the event area; the checklist of recommended equipment and supplies also includes maps and aerial photos. In any emergency response, the EOC will ask many *where* questions.

1. Where are the incidents? Where are they headed (e.g., wildfires, storms, chemical spills)?
2. Where are the people at risk?
3. Where are the health assets (fixed, such as hospitals and clinics)?
4. Where should we send our employees (e.g., during H1N1, the EOC at CDC in the USA coordinated the deployment of over 3000 CDC employees [50])?
5. Where are the mobile healthcare resources (e.g., ambulances) and other first responders?

Health departments have begun leveraging GIS as part of their EOCs, some developing stand-alone applications [51] and others using GIS in conjunction with products such as WebEOC.

Some data layers are event-specific, while others tend to be important regardless of the scenario. See Table 29.2 for a listing of essential and static data layers for a health EOC. Examples of data layers more specific to the type of event include first responder locations (e.g., real-time feeds re ambulance or helicopter locations), hospital diversion status, flood plain/zone boundaries, the extent of smoke plume (wildfires), electrical utility information, POD locations, etc.

Table 29.2 Essential and static data layers for a health EOC (Emergency Operations Center)

Essential data layers
• Event locations
• Hospitals and health facilities
• Public health departments
• Shelters
• Schools
• Roads
• Public transportation
• Restaurants and regulated facilities
• Major administrative boundaries
• Census data (especially demographic data, languages spoken, poverty, elderly, etc.)
• Post offices
• Municipalities
• Counties
• Surrounding state counties
Static health layers
• Advanced and basic life support
• Acute and nonacute care hospitals
• Points of distribution
• Blood banks
• RSS medical stockpile locations
• Health departments
• Hospitals from surrounding states
• Long-term care facilities
• Command centers
Real-time feeds (dynamic layers)
• Incidents/events
• Weather/traffic
• Medical stockpile
• SitStat surveys
• Hospital divert status
• Ambulance locations
GIS coordinators within EOCs perform essential functions, including the following
• Provide situational awareness to leadership and partners
• Map real-time spread of incidence/clustering
• Determine stockpile pod locations
• Locate and identify vulnerable populations
• Provide real-time mapping and analysis of vaccine inventory
• Monitor bed capacity/surge capacity in hospitals
• Help infected people locate treatment
• Facilitate mobile response and routing, especially in rural areas
• Identify gathering areas in high cluster areas
• Call up volunteers and staff by location
• Map distribution of care providers

Many health agencies have reinforced how important it is to be able to determine neighborhood-level population estimates during emergencies. It is worth noting that many of the *essential data layers* should not need to be collected from scratch . . . , rather, many are available already within the health department or from other governmental agencies. Some are likely to already be geocoded.

29.4.4 Community Health Assessment, Planning, and Profiling

Many health departments have used GIS to visualize the results of community health assessment activities. Community health assessment is a required activity for those local health departments seeking to be accredited in the USA. Moreover, hospitals are being asked to do more with community health assessments under the current health reform initiatives in the USA. More recently, health departments are recognizing the utility of GIS in many phases of the process. For example, a local health department in Texas has commented that *use of a GIS proved crucial in the planning, administration, and analysis of the community needs assessment* [52]. That same health department has suggested:

> The utilization of a GIS in orchestrating the community needs assessment efforts was essential in achieving the desired outcome of usable data.

Lifestyle was one of the four determinants of health referenced earlier in the Lalonde report and is an important component of community health. One of the ways that lifestyle is analyzed is through behavioral risk factor surveillance surveys. The CDC's Behavioral Risk Factor Surveillance System (BRFSS) was established in 1984. It is the world's largest continuously conducted telephone health surveillance system. BRFSS monitors state-level prevalence of the major behavioral risks among adults associated with premature morbidity and mortality. By 1994, all states, the District of Columbia, and three territories were participating in BRFSS. Many other countries have recognized the value of BRFSS and have asked CDC to help them establish and evaluate similar surveillance systems [56, 57].

Given the wide recognition of lifestyle as a determinant of health status, it is surprising there has not been more utilization of commercial lifestyle segmentation data by public health agencies and researchers. There have been some efforts in this area, but according to a GIS committee of the National Association of Central Cancer Registries [58]:

> Within public health, the roles of these lifestyle data are controversial. One of the problems with the use of these data are the lack of metadata about the procedures used to generate them (e.g., the data are developed using proprietary methods). Another problem is that the marketing terminology is perceived as politically incorrect (i.e., could not be used in a report to the public).

29.4.5 Environmental Health

Environmental health (EH) addresses all the physical, chemical, and biological factors external to a person, and all the related factors impacting behaviors. It encompasses the assessment and control of those environmental factors that can potentially affect health. It is targeted towards preventing disease and creating health-supportive environments. This definition excludes behavior not related to environment, as well as behavior related to the social and cultural environment, and genetics [59].

Environment was another of the four determinants of health referenced earlier in the Lalonde report. Environmental Health is one of the most well established areas of GIS utilization. Environmental health agencies at all levels of government and the partners that support them are increasingly using GIS technology to assess and protect the health of the populations they serve, understand the impacts of the environment on human health, and to improve environmental health services delivery. Environmental health organizations are interested in increasing their overall GIS capacity, so they may enhance environmental health practice both across programmatic areas (e.g., air pollution, water, toxics and waste, the built environment, etc.) and across common business functions, such as assessment, policy development, and assurance.

The National Environmental Public Health Performance Standards [60] from the CDC are an important benchmark for participating agencies to measure the capacity of their local environmental public health system or program. See Table 29.3 for GIS relevance to essential environmental public health service.

Recently, a number of countries initiated environmental public health tracking initiatives. CDC defines environmental public health tracking as the ongoing collection, integration, analysis, and interpretation of data about the following factors.

1. Environmental hazards
2. Exposure to environmental hazards
3. Health effects potentially related to exposure to environmental hazards [61].

Initiatives such as Environmental Public Health Tracking (EPHT) have resulted in not only national but also state-level EPHT portals [62], as well as the ongoing development of specific GIS extensions, such as the Rapid Inquiry Facility tool (RIF). RIF is an automated tool that provides an extension to Esri ArcGIS functions and uses both database and GIS technologies. Its purpose is to rapidly address epidemiological and public health questions using routinely collected health and population data. RIF was developed at the Imperial College London in collaboration with CDC [63]. More recently, there has been an emphasis on the built environment's impact on human health, including obesity and chronic diseases.

Table 29.3 GIS relevance to essential environmental public health services

Essential service	GIS relevance
1. Monitor environmental and health status to identify and solve community EH problems	GIS is a tool for EH assessment, analyzing trends, and communicating EH problems and risks to the public through static or interactive maps. GIS also has many functions helpful for exposure assessment, data aggregation, data management, and other linkages
2. Diagnose and investigate EH problems and health hazards in the community	GIS supports EH surveillance systems with more efficient data collection methodologies, better understanding of disease transmission dynamics, and a framework for outbreak investigation and response. There is universal consensus that GIS can be a useful aid at the beginning of an environmental epidemiology or risk assessment study [53]
3. Inform, educate, and empower people about EH issues	GIS facilitates targeting health communication geographically and demographically. Applications such as Community Commons (http://www.communitycommons.org/entities/92767d50-94a9-4bfd-acc6-946f0364e3e0) educate and empower people to understand EH issues
4. Mobilize community partnerships and actions to identify and solve EH problems	Maps are great tools for community engagement. Desktop GIS and web-based portals such as the ones listed above help mobilize community partnerships. Another example is the *rat information portal* in New York City (http://gis.nyc.gov/doitt/nycitymap/template?applicationName=DOH_RIP). GIS provides a framework for analyzing and solving many other EH problems (e.g., lead poisoning mitigation and prevention; integrated vector control to prevent malaria, dengue, etc.)
5. Develop policies and plans that support individual and community EH efforts	The quote *Documenting need is not enough; documenting where there is need is critical to intervention strategies* [54] holds true for EH practice. GIS has helped policymakers understand the scope of environmental health emergencies, the built environment, and the *zone of influence* of mobile sources of air pollution. GIS also plays a central role in public health impact assessments (https://ephtracking.cdc.gov/showHealthImpactDataGuide)
6. Enforce laws and regulations that protect EH and ensure safety	GIS-based methods help measure compliance with local laws (e.g., environmental setback regulations in jurisdictions) and compliance with spatial advertising restrictions in local and national laws (e.g., no tobacco advertising near schools). GIS-based methods are also utilized to geocode facilities and sites under regulation, route inspectors who regulate them, and track progress. GIS-based models allow planners to consider the safety of citizens when planning routes and testing preparedness plans
7. Link people to needed personal EH services and assure the provision of healthcare when otherwise unavailable	GIS helps identify underserved populations, barriers to service, and to coordinate service delivery among multiple agencies. GIS-enabled services locators help citizens understand what services are available in their area and which offices are responsible
8. Assure competent EH and personal healthcare workforce	Agencies and researchers have utilized GIS to assess workforce gaps in many different professions, including the EH workforce in California (163-page PDF at http://www.llu.edu/llu/sph/ophp/documents/eh_report2006.pdf) [55]. Geospatial analysis can characterize the pattern of deployment of the EH workforce and (with statistical modeling) analyze factors associated with the deployment pattern
9. Evaluate effectiveness, accessibility and quality of personal and population based EH services	GIS provides a framework for monitoring and evaluating programs and services. One of the most popular applications of GIS in health and human services is analyzing access to services
10. Research for new insights and innovative solutions to EH problems	GIS enables testing and considering options in both temporal and spatial contexts. Geospatial accuracy provides EH professionals and research partners with a more specific baseline for implementing and evaluating EH interventions and programs. GIS helps researchers aggregate data and understand complex, multidimensional relationships between pollution and disease

29.4.6 Chronic Disease Prevention and Control

GIS also supports numerous other chronic disease prevention and control activities. As mentioned throughout this chapter, GIS is utilized to support the work of cancer registries. GIS also helps support heart disease programs, stroke registries, diabetes registries [64], and even the siting of defibrillators. There has been extensive research into the impact of the built environment, including the *food environment* [65], on risk factors and health outcomes. Recently, CDC launched a chronic disease GIS exchange [66] that may highlight many of these efforts. It is also anticipated that GIS will contribute to the emerging field of *exposomics* [67].

Tobacco Control and Prevention

Tobacco control and prevention is a priority in many public health agencies, and GIS supports this winnable battle in numerous ways. GIS has been used to analyze tobacco advertising and compliance with the law [68–70], better understand local tobacco prevention efforts [71], prioritize communities for intervention [72], and model the impacts of tobacco taxes [73]. GIS has also been used to visualize policy efforts and policy changes in schools [74]. More recently, the New York City Health Department used GIS to monitor nicotine replacement therapy and found that the GIS analyses provided a unique, near real-time visual method of assessing participation patterns as well as the impact of media and outreach strategies [75].

Cancer Prevention and Control/Cancer Registries

Cancer and mapping have a long history together. Haviland's map of the distribution of cancer in British counties in 1892 reminds us that the cancer registries of today continue the long tradition in mapping cancer. Cancer registries were early adopters of GIS since the geocoding, data management, and spatial analysis capabilities of GIS benefit a wide range of cancer registry research questions and investigations.

29.4.7 Infectious Diseases

GIS for Pandemic Influenza and Epidemics

Early twenty-first century chapters in the history of mapping and GIS in public health include the Severe Acute Respiratory Syndrome (SARS) outbreak of 2003, the H1N1 pandemic of 2009, the Ebola epidemic of 2014, the Zika epidemic of 2015, and the COVID-19 Pandemic of 2020. During the SARS outbreak in 2003, many organizations used desktop and server GIS as a means of assembling and analyzing information on the spread and distribution of the disease. According to WHO, during outbreak response to SARS it used a custom-made geographical mapping technology to assist in the location of cases and rapid analysis of the epidemic's dynamics [76]. A number of organizations produced web-based interactive maps with SARS information at various levels of geography (see Fig. 29.4 for one example). Such mapping efforts educated the public (including travelers to potentially at-risk areas), assisted public health authorities in analyzing the spatial and temporal trends and patterns of SARS, and helped authorities assess and revise control measures [77].

Based on experiences with SARS and a recognition of the growing GIS capacities within governmental public health, many authors of national and subnational jurisdictions' pandemic influenza plans highlighted the importance of location-based information and real-time situational awareness during public health emergencies. Therefore, prior to the H1N1 pandemic of 2009 many preparedness plans included references to the need for GIS personnel in response efforts and considered ways GIS would be utilized to visualize and analyze incoming data in relation to key geographical information.

Planning and reality are often different. During the H1N1 and COVID-19 pandemic, GIS was helpful in more ways than public health planners had anticipated. At a high level, it helped visualize large amounts of rapidly changing data as the event progressed through time. At a local level, geocoding cases helped validate whether or not they were in a health department's jurisdiction [78].

In practice, health organizations used GIS to accomplish many functions. Table 29.4 is based on periodic interviews with GIS colleagues in state and local health departments in the USA.

HIV/AIDS

According to WHO and UNAIDS, over 33 million people around the world live with HIV. In addition to lives lost, there are many other substantial consequences of the HIV/AIDS epidemic, including orphans and vulnerable children need-

Fig. 29.4 In 2003, WHO and the Hong Kong Department of Health launched an interactive web mapping application to provide up-to-date, accurate information on the distribution of SARS in Hong Kong, China, and other parts of the world

Table 29.4 Pandemic influenza and GIS

Functional area	Practical applicability of GIS
Disease surveillance and cluster analysis	Geocoding cases of H1N1 for disease surveillance, disease diffusion, and cluster analysis. Maps of syndromic surveillance (e.g., citywide influenza-like-illness data from emergency departments). Producing maps and other outputs as situational awareness tools for managers, public information
Vaccines and antivirals	Health departments determined shipment sites for vaccines based on census demographics, used Thiessen polygons to determine antiviral apportionment, mapped the equitable distribution of vaccines based on general and high risk groups vs. actual distribution of vaccines, mapped antiviral distribution achieved through activation of stockpiles, monitored the availability of vaccines and antivirals geographically, instigated mass immunization outreach (geocoding and automated notification to under-represented areas), and created online vaccine facility locators
Situational awareness and decision support	Presenting maps to senior staff to apprise them of the changing situation (e.g., disease diffusion, school closures); where emergency rooms are located, where hospitalizations and deaths have occurred, determine populations in order to assist in calculating the number of supplies to be sent
Data verification and cleaning	Verifying postal codes, sorting by location, etc., in order to clean data going into numerous systems
Modeling and predictive analysis	Help predict where and how fast the pandemic will spread
Interactive mapping	Influenza vaccine locators and points of dispensing on existing interactive maps for the jurisdiction

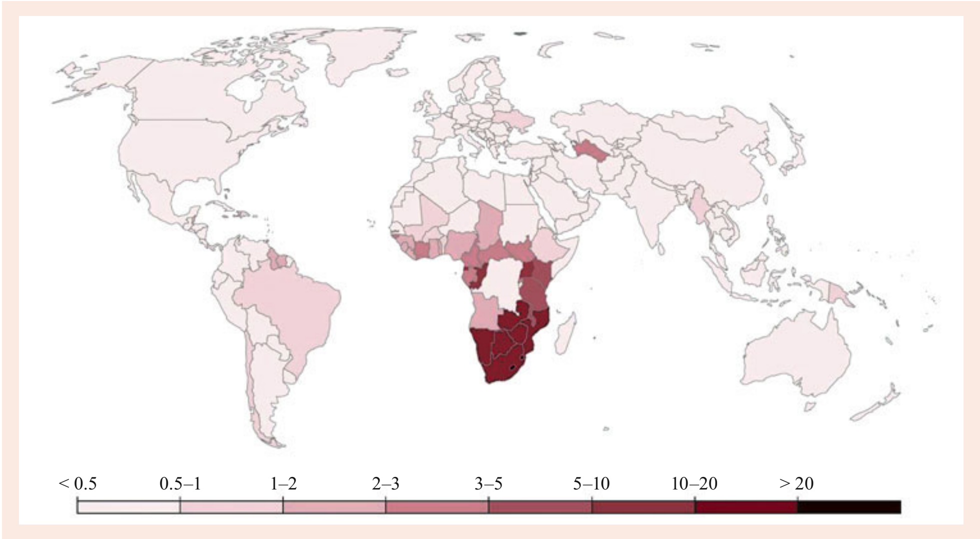


Fig. 29.5 2020: A global view of HIV Infection [79]

ing care, stresses on the health system, and lost economic productivity, among others. WHO suggests the HIV/AIDS epidemic still constitutes one of the greatest challenges to public health and international development [79].

According to [79], AIDS was the first major epidemic in recent world history that scientists were unable to follow in the spatial domain due to confidentiality issues. However, there have been numerous maps of HIV diffusion and in the early twenty-first century. GIS is an essential component of HIV/AIDS prevention and control. At the global level, HIV prevalence maps (such as the UNAIDS one, in Fig. 29.5) communicate the scope and distribution of HIV/AIDS worldwide and at the country level. Diffusion mapping has not only been local; it is now well known that major roadways have played a role in disease diffusion throughout many African countries.

The MEASURE (Monitoring and Evaluation to Assess and Use Results) evaluation project uses different strategies to collect and use data about health issues. For example, a tool for assessing and modifying HIV/AIDS prevention programs locally or nationally is the Priorities for Local AIDS Efforts (PLACE) method. PLACE can identify geographic areas that contain key HIV transmission networks.

The Spatial Data Repository (https://spatialdata.dhsprogram.com/home/) website has a collection of demographic and health surveys for many countries. GIS and GNSS help Ministries of Health, National AIDS Commissions (NACs), and their partners document the accessibility of services through nationwide health facility inventories. It also helps MOHs and NGOs understand the subnational distribution of population – and populations at risk – so that they may target resources effectively. This helps ramp up access to condoms, health communications, Voluntary Counseling and Testing services (VCT), Prevention of Mother-to-Child Transmission (PMTCT) services, and Antiretroviral Therapy (ART), among other things.

Malaria and Dengue

Public health professionals use GIS in a wide variety of ways to tackle mosquito-borne diseases such as malaria and dengue. GIS is an essential component of malaria prevention and control. According to *Hay* and *Snow*, *maps are essential for all aspects of the coordination of malaria control* [80]. The Global Malaria Action Plan recognizes the utility of mapping, especially for monitoring and evaluation [81]. The WHO regional dengue plan for 2008–2015 [82] includes a string of GIS-related items under Expected Result 10. In the WHO regional dengue plan, ministries of health are encouraged to conduct basic GIS workshops in 2009–2010 and to include GIS as part of their integrated vector management.

GIS enhances malaria and dengue surveillance and control at the national level all the way to the community level [83]. GIS and GNSS help NMCPs, dengue control programs, and NGOs understand subnational distribution of population – and populations at risk – so they may target resources effectively. This helps ramp up access to Insecticide-Treated Bednets (ITNs), indoor residual spraying campaigns (IRS), rapid diagnostics tests (RDTs),and artemisinin-based combination therapy (ACT), among other things. GIS has also enhanced ITN marketing and distribution efforts [84]. Recent reports and articles provide strong evidence that GIS enhances decision-making for national malaria control programs (NMCPs, for an overview of such activities in Zambias [85]) as well as national dengue control programs [86]. In some cases, vector control resources need to be targeted due to resource limitations. GIS helps here as well. Recent findings from Thailand suggest that integrated vector control programs using GIS-based foci (i.e., conducting community intervention campaigns within a defined radius of seropositive cases) are very effective [87]. Reference [87] also suggests that using a history of reporting dengue cases may be a practical tool for producing a GIS map of the risk areas, through which future vector control efforts could be targeted.

29.4.8 Animal and Veterinary Health

Cline notes [88]:

> while most pathogens transmitted in a human-to-human cycle are not constrained geographically, zoonotic and insect-transmitted diseases, in contrast, tend to be focal in distribution, with their maintenance cycles dependent upon exacting ecological conditions.

This dependency makes zoonotic and insect-transmitted disease well suited to GIS analysis. Over the last decade, GIS has become central to the work of many agricultural and veterinary health agencies. Agencies around the world have built GIS into applications supporting management of animal disease epidemics [89, 90]. Some organizations have also documented cost savings as a result of using GIS.

29.4.9 Human Services

Around the world, over half a million social workers and human service professionals help vulnerable and distressed people every day. Their clients include foster children, the elderly, the mentally ill, the homeless, the disabled, and many others in need. Social workers and human service professionals accomplish their work through hundreds of different programs at governmental agencies, nonprofit organizations, private practices, and other venues. These programs are in great demand due to the global economic crisis and numerous other drivers, including changing demographics. Increased demand for human services and social services brings new attention to concerns regarding access to services, efficiency (e.g., routing and logistics), and program integrity (e.g., fraud detection).

The adoption of GIS in the human services sector is similar to that in hospitals and public health. Human services organizations used GIS to understand the extent and scope of their market, where their clients originated, and where resources need to match service demand.

There is a long history of the helping professions utilizing maps for assessment, planning, and advocacy. *Hillier* [54] suggests that GIS can benefit social work by

1. Continuing and strengthening the social survey tradition
2. Providing a framework for understanding human behavior
3. Identifying community needs and assets
4. Improving the delivery of social services
5. Empowering communities and traditionally disenfranchised groups.

Hillier also argues [54]: documenting need is not enough; documenting *where* there is need is critical to intervention strategies.

If social work needs mapping, then human services agencies need GIS. In the late 1990s, *Wong* and *Hillier* [91] concluded that there were outstanding potential rewards for human service agencies using GIS in terms of agency planning, data analysis, policymaking, fundraising, client information, outreach, and other management and direct-practice functions. Over the last decade GIS has emerged as a critical tool for planning and understanding community needs, empowering citizens to locate services, enhancing the ability of helping professionals to provide information and referral; providing analytical capabilities to staff (e.g., policy questions, case management decisions, service eligibility), helping manage large amounts of program and administrative data, enabling logistics and program support (e.g., routing efficiency), and enhancing fraud detection. Professionals working in human services agencies ask many *where* questions such as

1. Where are the people who need help, those *at risk*?
2. Where are our services located?
3. Where should we send our employees today (e.g., many state agencies have staff numbering over 1000)?
4. What are the most efficient routes?
5. Where are the field staff located?
6. Where should we concentrate our limited fraud investigation capabilities?

One of the primary business functions of human service professionals is to provide case management for their clients whether in an agency office, at their clients' homes, or in other service delivery settings. Professional social work case managers are guided by standards [92] and routinely use computerized case management information systems. Such systems help track client history, progress toward meeting goals, and report results, among other functions. Case management systems with GIS at their core help human service professionals provide more effective and efficient case management. GIS leverages the power of place in establishing helping relationships, assessing complex problems, selecting problem-solving interventions, and helping clients to function effectively, all stated goals of case management. The *SchoolMinder* application in Illinois illustrates how adding server-based GIS tools to case management systems helps answer the *where* questions relevant to their cases and workflows. Since deploying the SchoolMinder application, the average distances for initial foster care placement in Cook County dropped from 15.9 to 4.0 km (median numbers dropped from 10.3 to 2.4 km) [93].

29.4.10 Hospitals and Health Systems

In recent years, GIS has expanded beyond the planning and marketing departments of hospitals to include a number of other applications beyond the analysis of hospital service areas and market share. At a regional and national level, there has been more interest in using GIS for service locators (consumer oriented) and health facility assessments (policy oriented). A number of organizations also partnered to come up with a standardized approach for identifying a health facility [94]. It is also worth noting that GIS can provide an inside view of the facility to determine when beds are available for cleaning after a patient has been discharged, where maintenance is needed on equipment, and the positioning of assets throughout the hospital. All of this can be managed and viewed through a web-based GIS tool. GIS can also help administrators and infection control professionals make better decisions about policies for the containment and immediate outbreak management response to secondary infections [95, 96].

29.5 GIS and HHS Education

The educational institutions that prepare the health and human services (HHS) workforce have recognized the demand for GIS-related courses and programs. A large number of accredited schools of public health are offering semester-long courses and summer institutes. Below are brief summaries of GIS within specific academic domains of public health.

29.5.1 Biostatistics and GIS

Biostatistics is the application of statistics to biological problems [97]. Since it often involves statistical analysis of large datasets, biostatistics benefits from GIS as a way of organizing large datasets (in a special context, by geographic identifiers) and also as a way of visualizing data for interpretation and hypothesis generation as an alternative to poring through spreadsheets or databases. One of the primary application areas of GIS to biostatistics is through cancer registries. It can be challenging to inform without misinforming, but researchers note that due to advances in the sciences of cartography, statistics, and visualization of spatial data they are constantly expanding the toolkit available to mapmakers to meet this challenge [98].

29

29.5.2 Community Health and GIS

Community health includes many programs such as maternal and child health, family health, health education, nutrition, and tobacco prevention and control. GIS adds value to community health in many ways, including

- Community health planning
- Analyzing health disparities
- Measuring access to health care and other services
- Understanding demographic data in relation to neighborhoods
- Mapping risk factors
- Analyzing high-risk locations (e.g., concentrations of liquor stores)
- Prioritizing interventions such as lead poisoning prevention
- Enforcing tobacco control legislation
- Using GIS for tobacco cessation promotion.

29.5.3 Epidemiology and GIS

Epidemiologists have often used maps to describe their study areas, aggregate data, link to other relevant datasets (e.g., demographic data), and conduct exploratory spatial data analysis. The enhanced time awareness of commercial GIS software will make the application of GIS in epidemiology grow even more. To understand the role of GIS in epidemiology, it is important to remember that epidemiology is largely quantitative but also observational. Epidemiologists explore patterns of disease across populations. GIS technologies are powerful and useful tools in observational epidemiology (with its focus on person, place, and time) as well as analytical epidemiology (identifying point sources, and controlling disease outbreaks, with its focus on location-based response activities).

Epidemiologists use the mapping and visualization functions of GIS to define study areas. They use analytic functions such as geocoding, buffering, exploratory spatial data analysis, field data collection (navigating to locations as well as recording location), and outbreak investigation. Location is proxy for exposure in environmental epidemiology.

29.5.4 Global Health and GIS

Global health priorities are largely focused on COVID-19, HIV/AIDS, malaria, maternal and child health, and preventing leading killers like diarrhea in children. However, many of these initiatives in global health have developed as vertical programs. Recently, donors have been calling for integrated approaches to combating disease and strengthening health systems. GIS has a special capacity to integrate and analyze data from a wide variety of programs that address malaria, tuberculosis, child and maternal health, clean water, food and nutrition, and education. GIS provides strategies for spatial data use for decision-making that support and strengthen linkages across health interventions towards the aim of overall health system strengthening. MOHs and the Nongovernmental Organizations (NGOs) supporting them use GIS in a broad range of programs and functional areas.

29.5.5 GIS and e-Health

There are a number of definitions for e-Health. WHO defines e-Health as the use of information and communication technologies (ICT) for health [99]. HIMSS (Healthcare Information and Management Systems Society), a not-for-profit membership organization devoted to healthcare transformation through the effective use of health information technology, defines e-Health as [100]

> The application of Internet and other related technologies in the healthcare industry to improve the access, efficiency, effectiveness, and quality of clinical and business processes utilized by healthcare organizations, practitioners, patients, and consumers to improve the health status of patients.

Since GIS is information technology (IT), it is covered by the above-referenced definitions for e-Health. Below are several examples of how GIS enhances e-Health.

The geographic information contained in electronic health records and health information systems is a critical component for the detection of disease outbreaks. More specific geographic information in health records increases options for detecting outbreaks (i.e., one can always aggregate up but not down). Specific geographic information within health information systems also helps health officials determine the extent of an outbreak. With such knowledge, health organizations may use GIS to determine what resources are available in close proximity (e.g., facility diversion status, stockpile locations, volunteers). Such analysis is essential for response and communication with the public. Surveillance is not just important for infectious/communicable diseases; chronic disease registries (cancer, stroke, etc.) also benefit from more specific geographic information. Specific and standardized location information embedded in electronic health records (and registries) facilitates more targeted spatial analyses across health information exchanges [101, 102].

GIS helps deliver the e-Health promises of fostering participation by consumers, as well as addressing disparities (e.g., in access to services and health outcomes). The strategic planning documents of numerous e-Health initiatives suggest e-Health will foster participation by consumers by allowing them to better manage their own care and be more informed in decision-making. If this is to be the case, then consumers need to know where services are located. Also, health authorities and community-based organizations need to understand where the consumers needing various services are located and what disparities currently exist in not only their access to services but also health outcomes and health risk factors. Such location data is not just for research, it has extreme practical value. The geocoordinates of health facilities are necessary for analyzing access to services but may also feed public-facing services locators. Moving forward, it is anticipated that more consumers will utilize such services locators from smartphones and other mobile devices.

GIS helps us understand unwarranted geographic variation in health services delivery, supporting the e-Health goal of supporting providers in the delivery of safer, more effective, and more efficient healthcare. Numerous researchers have documented the unwarranted geographic variation in health services delivery [22], which not only has an impact on costs to consumers and governments but also on health outcomes. Many e-Health initiatives facilitate the increased availability of health services and health outcomes data. In helping place such data on the map, GIS provides new eyes to researchers, clinicians, and practitioners to understand and address unwarranted geographical variations.

GIS helps deliver e-Health's promises for more transparency, i.e., timely, accurate, and comprehensive reporting on health system activities and outcomes. MOHs, subnational health departments, and NGOs are on the front lines implementing health improvement initiatives around the world, including chronic disease prevention and control; HIV/AIDS prevention, care and treatment; malaria prevention and control; and maternal and child health, among others. Elected representatives, governing bodies, taxpayers, and external donors want the health organizations they fund to design and implement successful programs. Health organizations (both governmental and nongovernmental) also have increasing responsibilities to partner with other government agencies and NGOs (e.g., joint proposals), coordinate across vertical programs, and build local capacity. Therefore, they need tools that help them report on what they are doing [103] and collaborate with others. Health organizations are also moving from paper processes to using tablets and mobile phones for data collection [104].

The WHO Health Metrics Network has emphasized how important it is to map human resources, budgets, and expenditures at the national and district levels [105]. There is an increasing number of ArcGIS server applications that are examples of providing this type of transparency. Moving forward, it is anticipated that many MOHs, subnational health departments, and NGOs will map various health management information system (HMIS) data (whether it be operations data or health outcomes).

The personalization of health is not new. Medicine has always had a person at the center of its inquiry and practice. The Hippocratic oath has always had a single person in mind with its policy of do no harm. However, studying one person at a time is labor intensive and does not produce the generalizability that is required to impact large groups of people. This, of course, is where public health comes in by providing the broad brush approach to health problems and then raising awareness as to what requires our attention as a society. While the public health disciplines have studied population health at group levels, they seldom focus on individual levels. Geomedicine, on the other hand, provides a different type of analysis: it attempts to link the highly governable health information to the unique context of the individual. It proposes that factors in our environment have a substantial impact on our own personal health.

Geomedicine attempts to harness the power of a GIS to present environmental context in a scalable fashion, where the knowledge of medicine can impact how humans choose to engage with the underlying context. To assure that both the information upon which geomedicine is used is useful to both physicians and patients, all health data must be geographically accurate, and accuracy is solely a function of how well computerized patient registration systems collect, manage, and maintain highly granular geographic data in a timely manner. In Sect. 29.6 we will briefly review the relevant history of Geomedicine and describe where its promise lies in medical care.

29.6 An Abbreviated History of "Geomedicine"

Hippocrates medical teachings most likely inspired all those persons who have framed arguments for the use of the following terms: geomedicine, medical geography, geospatial medicine, geographical geology, and environmental geology. Amy Blatt's recent book, *Health, Science and Place* provides an overview of geography in medicine [106].

Many of these earlier authors could have not imagined the implications of big data and the powerful analytical capacities of modern information technologies and what they could offer to those studying human health from a geographic perspective. Another excellent reference describing some of the early ways that geography and medicine have intersected can be reviewed in Koch's book *Cartographies of Disease – Maps, Mapping and Medicine* [107]. One of the most popular environmental geological textbooks *Environmental Geology* [108] by *Montgomery*, an environmental geologist, devotes an entire chapter to the topic of geomedicine, or as she describes it the "science of geomedicine", presents much of the extant research in the various fields of geology.

In contrast to the extensive basic research conducted by many academics, there have also been significant contributions by several health practitioners and researchers whose aim was to improve human health in a more direct, individual way. These persons did not start out to prove a causative relationship between what they observed, yet they inspired many others to undertake that work. Many of these researchers were conducting exploratory research and in-depth descriptive research, in search of clues to what they thought were obvious problems and opportunities related to personal health. We briefly introduce some of these applied researchers to add context to the discussion of geomedicine.

29.6.1 Applied Geomedicine: An Early Blueprint

In 1918, George Palmer, a practicing physician became Superintendent of the Springfield Illinois Department of Health and launched an ambitious survey of the sanitary systems of that city of about 80 000 people. Palmer, a trained physician, conducted his community survey along the lines he was trained in as a physician (e.g., comprehensive diagnosis, testing and verification, treatment and follow-up). Palmer's research considered a person's entire living situation, his work, his genetics, his worries, his families, his habits of daily living, all part of his inquiry. Upon completing his work, he was asked to prepare a paper for the annual meeting of the American Public Health Association. The paper he delivered, "*The Diagnosis of the Sick City*" [109] could be described as a blueprint or framework that cities across America would try to emulate even today. Clearly his approach, if implemented in the same way today, would be an expensive and time-consuming way to inform an individual health record with relevant contextual information for every patient.

29.6.2 Applied Geomedicine: Understanding Why Place Matters

In 1973, John Wennberg, a physician researcher published one of the landmark studies of geographic variation of health care practice and delivery. His lifetime research on geographic variation is chronicled in an excellent overview of his 40+ year career [110] in a magazine article published by the Dartmouth Medical School and is worth a careful read for anyone with similar ambitions to change the way medical care is practiced. Wennbergs's research began one of the most prolific periods of exhaustive and in-depth research into the way healthcare is practiced, delivered, and financed geographically. Wennberg had many things going for him: a medical degree from Stanford University, a public health degree from Johns Hopkins University, a teaching position at Dartmouth College and a staff of experienced and seasoned medical researchers, a GIS specialist, and confidential access to the entire national Medicare database of patient records.

Wennberg also had the attention of anybody in Washington, DC, who was concerned about the way the nations Medicare program was evolving in terms of cost and quality of medical care. Much has been written about Wennberg's work in other places, including descriptions of his Dartmouth Atlas projects referenced earlier in this chapter. Geographic information systems became an integral part of every research study conducted by Wennberg's Medicare research program and the work of the Center for The Evaluative Clinical Sciences at Dartmouth College and is one of the best examples of using geography and GIS in a meaningful and persistent way to help society understand the outcomes of the Medicare health care on patients and healthcare providers.

29.6.3 Applied Geomedicine: Its Value to Physicians and Patients

In 2008, William Davenhall was invited to give a presentation to TEDMed [111] that would explore how geographic information could help people live healthily. His presentation to largely a medical audience took many by surprise, as many in the audience had never heard of geographic information systems nor had they given much thought to how it might be used in the practice of medicine or in personal health.

Davenhall made the case for the smarter use of geographic information and introduced the idea of creating "medical place histories" inside a person's electronic medical record that could be used to provide physicians with greater health-relevant knowledge about the patient's present and historical environmental exposures. The 8-min presentation challenged the audience to consider ways that medical education could begin to teach physicians about the value of geographically relevant health information in their medical practice. His message was extensively reported by the major medical media, including *Scientific American* [112], *Huffington Post* [113], *Information Week* [114], and *WIRED* [115].

In the years following the TEDMed presentation, Esri (Davenhall's employer at the time) published *Geomedicine: Geography and Personal Health* [116], and developed a smartphone application called "My Place History". The app was initially an iOS app (and later as a PC web application) that could be downloaded onto personal smartphones. The app allowed users to inventory all places they had ever lived (using street addresses), assuring geographical accuracy. The app then compared each address entered to a large governmental reference database that has inventoried all the toxic release sites (by latitude and longitude) across the US since 1987 [117]. The data included in this national database contains all locations (places) where a list of specific human cancer-producing toxic materials were buried, injected into soil, or released into the atmosphere. The app then allowed the user to print and share a report and map showing the proximity of places they had lived, worked, or played in location to the toxic release inventory sites.

A critical review of geomedicine and Davenhall's suggested "innovations" is extensively covered in a paper published in the *International Journal of Health Geographics* [118]. Clearly, the easy availability of the TEDMed presentation, eBook, and a downloadable working demonstration app has contributed to the wide diffusion of published materials to students in the health and human services, medical informatics, and computer science fields around the world. An 8-min story, well told, accompanied by a digital eBook and a smartphone app greatly accelerated the introduction of the applied geomedicine idea.

29.6.4 Applied Geomedicine: Early Evidence of Adoption

In 2014, Duke University Center for Health Informatics held the first Geomedicine Summit for health systems. Several large hospital-based health systems presented their work in the integration of massive amounts of geo-tagged patient and health care delivery data into their institutional information systems on a daily basis. Some applications were more focused on improving the use of geographically relevant data within the physician's office encounter while others enabled real-time geocoding into their patient registration system.

As the capabilities of GIS technologies and big data begin to advance, organizational capabilities to integrate large amounts of widely different types of health-relevant data that is geographically specific, will speed adoption of GIS tools and technologies. The demand for a more inclusive data model for viewing the "whole patient" will become more popular, and geomedicine is expected to help accelerate the "why" part of GIS adoption. Clearly, there are some major hurdles to overcome in making geographic information systems part of the institutional approach to generating useful knowledge from their existing, and very expensive data generation systems – like patient registration or medical records. Listed here are several large hospital-based health systems that have taken steps toward greater systematic use of GIS across their entire operational environment: Baystate Medical Center, Springfield MA, Loma Linda University Medical Center, Loma Linda, CA, the University of Kentucky Medical Center, Lexington, KY, Duke Medical Center, Durham, NC, and the Children's National Hospital, Washington, DC [119]. Geomedicine, as an applied operational strategy to deliver patient-centric care could provide a sound and meaningful justification for its adoption.

29.6.5 The Future of Applied Geomedicine

Applied means "finding a solution for an immediate problem facing society" [120]. The "problem" is the massive underutilization of geographically health-relevant data in the doctor's examination room. The "solution" is to build an information system that assures that no relevant piece of a patient's health "puzzle" is unintentionally overlooked or goes unused in the assessment of their illness, symptoms, disease, treatment, or healing process.

There is adequate peer-reviewed medical research that can support applied geomedicine, or at least make a considerable contribution to endeavor to meet human health needs. As the sheer volume of geographically tagged health data (geotagged) becomes prolific, yet this data largely goes underused to help patients with their health issues and concerns. Change will happen, and a new way to look at the data will emerge. As higher volumes of medical insurance claims generated daily from physicians' offices, hospitals, labs, and pharmacies across the US fill up cloud-based data silos, the imperative to make this data do something good for patients and health-seeking consumers will likely appear. People's care systems and their respective "providers" will finally seriously engage with GIS technologists who have the tools to help them find cost effective ways to implement advanced GIS systems capable of managing high volume transactional health and social data with accurate geographic granularity and time sensitivity. Applied geomedicine could easily become the next great "medical innovation".

29.6.6 COVID-19 Pandemic and GIS

The COVID-19 pandemic of 2020 created the environment in which GIS tools, techniques, and digital data management were essential. Understanding the spatial-temporal dynamics of COVID-19 has been critical to its mitigation [121]. Attractive interactive web sites, world-wide, got underway exceptionally fast. One example is the COVID-19 Dashboard, developed by the Center for Systems Science and Engineering at Johns Hopkins University [122]. Data analytics took on a new sense of life saving information. Unprecedented reporting of everything "COVID-19" greatly expanded world-wide recognition of geographical information system technology and its applications. In the COVID-19 response, some public health agencies have utilized GIS to bring additional attention to issues of equity, e.g., in vaccine distribution and administration. The CDC's Vaccination Equity Map is based on the percentage of population fully vaccinated, as well as the Social Vulnerability Index (which uses Census data on categories like poverty, housing, and vehicle access) [123].

Several major outcomes and challenges for GIS in the COVID-19 response and recovery have been observed and include:

Greater Need for Accelerated Analytical, Reporting, and Mapping

Numerous public health, geographic information specialist, and data scientists began to develop analytical software, maps, and dashboards to monitor the spread and subsequently, the immunization activities that followed after vaccines were approved for distribution. The quick response of the information specialists within public health departments across the United States, along with hundreds of health authorities and private public health research organizations in almost every national health ministry began to use their analytical tools, especially GIS to their advantage in the tracking of the disease spread, reporting of cases, deaths, and the availability of personal protective equipment (PPE). Many of these activities will demand even greater speed and scalability.[1]

[1] Source: Research notes prepared by William Davenhall, email: bill.davenhall@gmail.com, 2020–2021

29

Increased Public Use of Highly Granular Geographic Health Data

A useful and powerful feature of using GIS tools is its ability to create near real-time visualization of relevant case data. Geographic information systems were also used to make assessments of the operational activity of various vaccine transportation and distribution activities. Consumers had the ability to locate vaccination sites and even schedule appointments to receive the COVID-19 vaccine. The need for more accurate field data became obvious and will most likely evolve into a political imperative. The presence of many non-health professionals communicating with the general public using geographical information system applications has probably never been greater than during the COVID pandemic.[2]

From public health organizations and their various governmental sponsors, to private corporations making plans for worker safety and business continuance, the impact of the pandemic will be likely be viewed as a "watershed" moment that demonstrated the unequivocable critical role that geographic information system technology and software applications played in delivering actionable and timely information when it was most critically needed – everywhere.[3]

29.7 Summary

GIS continues to find its place in health and human services (HHS). Numerous research projects explore novel uses, while over time, many GIS functions and operations have become standard practice within health and human service organizations. The value of GIS in all health and social care activities continues to increase. As more nations face uncertain economic and political times, these fundamental human service market segments must get greater value out of their transactional data for use in service delivery, as well as to improve the effectiveness and accuracy of their analytical endeavors.

The future of the use of geographic information systems technology in health and human services is extremely bright. The increasing focus on electronic health records, and the corresponding geographic information contained within them, will open up many new possibilities for population health analysis and planning. GIS has been found to be an essential technology in a wide variety of health and human services agencies and activities. Over 1,000 case studies have been identified by just one GIS company. These case studies span the continuum of HHS organizations.

The advent of cloud computing will increasingly make GIS more affordable for every MOH, while allowing software developers to focus on creating custom, easy-to-use applications to serve specific needs while maintaining a framework to seamlessly share relevant information across the entire health and human services continuum. Location will be an integral part of every health and human service.

[2] Source: Research notes prepared by William Davenhall, email: bill.davenhall@gmail.com, 2020–2021

[3] Reference: (Quote by Bill Davenhall, Geomedicine Analyst, Davenhall Associates, LLC, Unpublished Memorandums, 2020–2021) eMail: bill.davenhall@gmail.com

References

1. WHO: Constitution of the World Health Organization, New York, 1946. https://www.who.int/governance/eb/who_constitution_en.pdf. signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948
2. WHO: International Classification of Diseases (ICD). WHO, Geneva (2011). available at http://www.who.int/classifications/icd/en/
3. WHO: History of the development of the ICD. WHO, Geneva (2011). available at http://www.who.int/classifications/icd/en/HistoryOfICD.pdf
4. Lalonde, M.: A New Perspective on the Health of Canadians: A Working Document. Government of Canada, Ottawa (1974). available at https://nccdh.ca/resources/entry/new-perspective-on-the-health-of-canadians
5. Winslow, C.E.A.: The untilled field of public health. Mod. Med. **2**, 183–191 (1920)
6. Merriam-Webster's Online Dictionary: Geography (2010). http://www.merriam-webster.com/dictionary/geography
7. Encyclopædia Britannica: Medical geography (2010). http://www.britannica.com/EBchecked/topic/372244/medical-geography0
8. Oliver, M.: Editorial. J. Am. Board Fam. Med. **23**(1), 1–3 (2010)
9. Hippocrates: On airs, waters, and places (2000). http://classics.mit.edu/Hippocrates/airwatpl.mb.txt. The Internet Classics Archive, ed. by D.C. Stevenson
10. Meade, M.S., Earickson, R.J.: Medical Geography. Guilford, New York (2000)
11. Koch, T.: Cartographies of Disease: Maps, Mapping and Medicine. ESRI, Redlands (2005)
12. Finke, L.L.: Versuch einer allgemeinen medicinisch-praktischen Geographie, worin der historische Theil der einheimischen Völker- und Staaten-Arzneykunde vorgetragen wird. Weidmannsche Buchhandlung, Leipzig (1792)
13. Frerichs, R.R.: John Snow (2017). http://www.ph.ucla.edu/epi/snow.html
14. Davenhall, W.F.: An ESRI Internal Research Project (2002)
15. Cline, B.L.: New eyes for epidemiologists: aerial photography and other remote sensing techniques. Am. J. Epidemiol. **92**(2), 85–89 (1970)
16. Audy, J.R.: A summary, topographical account of scrub typhus 1908–1946. Studies in the distribution and topography of scrub typhus. Bull. Inst. Med. Res. Kuala Lumpur **1**, 84 (1949)
17. Washino, R.K., Wood, B.L.: Application of remote sensing to vector arthropod surveillance and control. Am. J. Trop. Med. Hyg. **50**, 134–144 (1994)
18. Davenhall, W.F.: Unpublished client research report. Spindletop Research, Lexington (1973)
19. Henriques, W., Lee, C., Howie, M.: Site community health investigation using GIS. In: ESRI 1999 User Conf. Proc. ESRI, Redlands (1999). http://proceedings.esri.com/library/userconf/proc99/proceed/abstracts/a452.htm

20. Rosner, B.: Fundamentals of Biostatistics, 7th edn. Daxbury, Pacific Grove (2010)
21. Wennberg, J., Gittelsohn, A.: Small area variation in health care delivery. Science **182**, 1102–1108 (1973)
22. Goodman, M.: Geography is destiny: The epidemiology of health care. In: ESRI 2009 Health GIS Conf. Proc. ESRI, Redlands (2009). http://proceedings.esri.com/library/userconf/health09/docs/plenary/goodman.pdf
23. Dartmouth Institute: The Dartmouth Atlas (2011). www.dartmouthatlas.org. Dartmouth Institute for Health Policy and Clinical Practice, Lebanon
24. Bernal-Delgado, E.: ECHO project (2011). http://www.echo-health.eu. Institute for Health Sciences Aragon, Zaragoza
25. Garb, J.L., Ganai, S., Skinner, R., Boyd, C.S., Wait, R.B.: Using GIS for spatial analysis of rectal lesions in the human body. Int. J. Health Geogr. **6**, 11 (2007)
26. HuBMAP Consortium (2019): The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. Nature **574**, 187–192 (2019)
27. Linger, S., Wolinsky, M.: Geophylogeny: Merging views of evolution and migration (2006). https://www.researchgate.net/publication/267027475_Geophylogeny_Merging_Views_of_Evolution_and_Migration
28. Miranda, M.L., Silva, J.M., Overstreet Galeano, M.A., Brown, J.P., Campbell, D.S., Coley, E., Cowan, C.S., Harvell, D., Lassiter, J., Parks, J.L., Sandel, W.: Building geographic information systems capacity in local health departments: Lessons learned from a North Carolina project. Am. J. Public Health **95**, 2180–2185 (2005)
29. WHO GIS Centre for Health: Timely and reliable decisions save lives (2021). https://www.who.int/data/GIS. WHO, Geneva
30. WHO: Microplanning for immunization service delivery using the Reaching Every District (RED) strategy (2009). http://whqlibdoc.who.int/hq/2009/WHO_IVB_09.11_eng.pdf. WHO, Geneva
31. California Hospital Adverse Events Map 2013–2019: https://www.arcgis.com/apps/webappviewer/index.html?id=3968945bebb047b085142df9f06b0dfC. California Department of Public Health, Sacremento
32. WHO: Intercountry consultation for developing a regional strategy for patient safety. Kuwait City, Kuwait (2004) (2005). https://apps.who.int/iris/bitstream/handle/10665/255388/WHO_EM_HCD_053_E_en.pdf?sequence=1&isAllowed=y. WHO EMRO, Cairo
33. Ali, D., Levin, A., Abdulkarim, M., Tijjani, U., Ahmed, B., Namalam, F., Oyewole, F., Dougherty, L.: A cost-effectiveness analysis of traditional and geographic information system-supported microplanning approaches for routine immunization program management in northern Nigeria, Vaccine https://doi.org/10.1016/j.vaccine.2019.12.002
34. Hamblin, K.: Transport network optimization: Features in practice. Usaid Deliv. Proj. Supply Chain Manag. E-Newsl. Q. **2010**(2), 7–9 (2010)
35. Mashal, T., Nakamura, K., Kizuki, M., Seino, K., Takano, T.: Impact of conflict on infant immunisation coverage in Afghanistan: A countrywide study 2000–2003. Int. J. Health Geogr. **6**, 23 (2007)
36. UNICEF: Improving Vaccination Coverage and Reducing Inequities: Use of GIS in Immunization Programs. FINAL REPORT 6 February 2017. https://www.unicef.org/media/58031/file. UNICEF, New York
37. Urquhart, G.A., Williams, W., Tobias, J., Welch, F.J.: Immunization information systems use during a public health emergency in the United States. J. Public Health Manag. Pract. **13**(5), 481–485 (2007)
38. ArcGIS: Health and human services tracks stockpile shipments on the web. Arcnews online (2006). http://www.esri.com/news/arcnews/fall06articles/health-and-human.html
39. Brems, B.: GIS analysis utilized for decision-making during a mass vaccination. In: ESRI 2007 Health GIS Conf. Proc. ESRI, Redlands (2007). http://proceedings.esri.com/library/userconf/health07/docs/gis_analysis.pdf
40. Kittayapong, P., Yoksan, S., Chansang, U., Chansang, C., Bhumiratana, A.: Suppression of dengue transmission by application of integrated vector control strategies at sero-positive GIS-based foci. Am. J. Trop. Med. Hyg. **78**(1), 70–76 (2008)
41. Lester, G.: The threat of avian flu in Pennsylvania. Penn News (2005). https://news.psu.edu/story/457014/2017/03/21/avian-flu-threat-re-emerges-new-outbreaks-are-reported-us-flocks. Univ. Pennsylvania, Philadelphia
42. Johnson, M., Wilfert, R.A.: Rapid needs assessment and GIS. Focus Field Epidemiol. **5**(3), 1–5 (2010)
43. Horney, J.A., Moore, Z., Davis, M., MacDonald, P.D.M.: Intent to receive pandemic influenza A (H1N1) vaccine, compliance with social distancing and sources of information in NC. PLoS One **5**(6), e11226 (2009)
44. Henning, K.J.: Overview of syndromic surveillance: What is syndromic surveillance. MMWR Morb. Mortal. Wkly. Rep. **53**(Suppl), 5–11 (2004)
45. Nelson, C., Lurie, N., Wasserman, J., Zakowski, S., Leuschner, K.: Conceptualizing and Defining Public Health Emergency Preparedness. Rand Working Paper, vol. WR-543. Rand Corp, Santa Monica (2008). available at http://www.rand.org/pubs/working_papers/WR543.html
46. Holt, J., Mokdad, A., Ford, E., Simoes, E., Mensah, G., Bartoli, W.: Use of BRFSS data and GIS technology for rapid public health response during natural disasters. Prev. Chronic Dis. **5**(3), A97 (2008)
47. Johnson, E.: GIS assists DC FEMS to attain an unprecedented service level. In: ESRI 2007 Health GIS Conf. Proc. ESRI, Redlands (2007). http://proceedings.esri.com/library/userconf/health07/docs/gis_assists.pdf
48. Khan, S., Grigsby, S.F., Moore, P., Bagwell, D.A., Tao, G., Faren, S., Melton, M., Liem-McFadden, M., Morrow, A., Fernandez, R.: Geographic information systems: Using technology for public health disaster planning. In: Am. Public Health Assoc. 2006 Annu. Meet. (APHA), vol. 130351. (2006). http://apha.confex.com/apha/134am/techprogram/paper_130351.htm
49. WHO WPRO: Managing Health Emergencies: A Guide for Establishing, Operating and Evaluating an Emergency Operations Centre. WHO, Manila (2006)
50. CDC: The 2009 H1N1 Pandemic: Summary highlights, April 2009–April 2010. Atlanta. https://www.cdc.gov/h1n1flu/cdcresponse.htm
51. Werbel, R.: Hippocrates, New Jersey's health system. In: ESRI Health GIS Conf. Proc. ESRI, Redlands (2007). http://proceedings.esri.com/library/userconf/health07/docs/hippocrates_new_jerseys.pdf
52. Kazda, M.J., Beel, E.R., Villegas, D., Martinez, J.G., Patel, N., Migala, W.: Methodological complexities and the use of GIS in conducting a community needs assessment of a large US municipality. J. Community Health **34**(3), 210–215 (2009)
53. Betts, K.: Mapping the environment. Environ. Health Perspect. **105**(6), 105 (1997)
54. Hillier, A.: Why social work needs mapping. J. Soc. Work Educ. **43**(2), 205–221 (2007)
55. Dyjack, D.T., Case, P., Marlow, H., Soret, S., Montgomery, S.: California's county and city environmental health services delivery system. J. Environ. Health **69**(8), 35–43 (2007)
56. https://www.cdc.gov/brfss/about/about_brfss.htm

57. National Center for Chronic Disease Prevention and Health Promotion: Behavioral risk factor surveillance system (2019). http://www.cdc.gov/chronicdisease/resources/publications/aag/brfss.htm
58. Wiggins, L. (ed.): Using Geographic Information Systems Technology in the Collection, Analysis, and Presentation of Cancer Registry Data: A Handbook of Basic Practices. North American Association of Central Cancer Registries, Springfield (2002)
59. WHO: Environmental health (2011). http://www.who.int/topics/environmental_health/en/. WHO, Geneva
60. Colon, J., Li, J., Elligers, A.: Environmental Public Health Performance Standards (EnvPHPS). NACCHO, Washington (2011)
61. CDC: National Environmental Public Health Tracking Program: Glossary. Centers for Disease Control and Prevention, Atlanta (2009)
62. Kassinger, C.: The role of GIS in the environmental public health tracking network. In: ESRI 2010 Fed. User Conf. Proc. (2010). https://proceedings.esri.com/library/userconf/feduc10/papers/user/craig_kassinger.pdf
63. Beale, L., Abellan, J.J., Hodgson, S., Jarup, L.: Methodologic issues and approaches to spatial epidemiology. Environ. Health Perspect. **116**, 1105–1110 (2008)
64. Subhani, S.: Use of geographical information system within web-based Saudi National Diabetes Registry. In: 6th Natl. GIS Symp., Saudi Arabia (2010). http://www.saudigis.org/FCKFiles/File/SaudiGISArchive/5thGIS/Papers/T3_4.doc
65. Gustafson, A., Duffey, K., Ammerman, A., Creamer, N., Worthington, S., Ickes, S.: Understanding the local food environment through a community food system assessment. https://www.researchgate.net/publication/266779163_Understanding_the_local_food_environment_through_a_community_food_system_assessment
66. CDC: Chronic Disease Geographic Mapping. Centers for Disease Control and Prevention, Atlanta (2010)
67. CDC: Exposome and exposomics (2010). http://www.cdc.gov/niosh/topics/exposome/. Centers for Disease Control and Prevention, Atlant
68. Scott, M.M., Cohen, D.A., Schonlau, M., Farley, T.A., Bluthenthal, R.N.: Alcohol and tobacco marketing: Evaluating compliance with outdoor advertising guidelines. Am. J. Prev. Med. **35**, 203–209 (2008)
69. Hernández-Ávila, J., Tirado-Ramírez, E., Santos-Luna, R., Reynales-Shigematsu, L.M.: Use of geographical information systems for billboards and points-of-sale surveillance in two Mexico cities. Salud Publica Mex. **49**(Supl. 2), 241–246 (2007)
70. Reynales-Shigematsu, L.: Surveillance of tobacco advertising and products on point of purchase marketing in two cities in Mexico using a geographic information system (GIS). In: Salud Publica Mex. **49**(Supl. 2), Cuernavaca ene (2007). http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0036-36342007000800014
71. Goetze, D.: Using local data and GIS to analyze county level comprehensive tobacco prevention in Indiana. In: Am. Public Health Assoc. 2007 Ann. Meet. (2007). https://apha.confex.com/apha/134am/techprogram/paper_138577.htm
72. Li, W., Land, T., Zhang, Z., Keithly, L., Kelsey, J.L.: Small-area estimation and prioritizing communities for tobacco control efforts in Massachusetts. Am. J. Public Health **99**(3), 470–479 (2009)
73. Burgess, P.: San Bernardino County, California, agency analyzes tax impact with GIS. ArcNews (2005). http://www.esri.com/news/arcnews/spring05articles/sanbernardino-county.html
74. Morris, J.: Tobacco-free school policy adoption in Oklahoma. In: Am. Public Health Assoc. 2008 Annu. Meet. (2008). https://www.researchgate.net/publication/266819165_Tobacco-free_school_policy_adoption_in_Oklahoma
75. Czarnecki, K.D., Goranson, C., Ellis, J.A., Vichinsky, L.E., Coady, M.H., Perl, S.B.: Using geographic information system analyses to monitor large-scale distribution of nicotine replacement therapy in New York City. Prev. Med. **50**(5/6), 288–296 (2010)
76. WHO: Communicable disease surveillance and response severe acute respiratory syndrome (SARS): Status of the outbreak and lessons for the immediate future (2003). http://www.who.int/csr/media/sars_wha.pdf. WHO, Geneva
77. Boulos, M.: Descriptive review of geographic mapping of severe acute respiratory syndrome (SARS) on the Internet. Int. J. Health Geogr. **3**, 2 (2004)
78. Goranson, C.: Private communication. CDC Analysis-Visualization-Reporting webinar. (2010)
79. Joint United Nations Programme on HIV/AIDS (UNAIDS): Confronting inequalities: Lessons for pandemic responses from 40 years of AIDS. https://www.unaids.org/en/resources/documents/2021/2021-global-aids-update
80. Hay, S.I., Snow, R.W.: The Malaria Atlas Project: Developing global maps of malaria risk. PLoS Med. **3**, e473 (2006)
81. Roll Back Malaria Partnership: The global malaria action plan (2008). https://www.unhcr.org/4afac5629.pdf. Boston Consulting Group, Boston
82. WHO SEARO: Asia-Pacific dengue strategic plan (2008–2015) (2008). https://apps.who.int/iris/bitstream/handle/10665/128710/Sea_rc61_11%20Inf.%20Doc%20done.pdf?sequence=2&isAllowed=y. WHO South East Asia Regional Office, New Delhi
83. Amoya, F.: Locating the cases: A community GIS-based malaria information surveillance system. In: ESRI 2008 Health GIS Conf. Proc. ESRI, Redlands (2008)
84. Chacha, P.: GIS support on fighting malaria. In: ESRI Eastern Africa 2008 User Conf. Proc. ESRI, Redlands (2008)
85. NMCC: Integrating Geographic Information Systems (GIS) into Indoor Residual Spraying (IRS) Planning and Implementation. National Malaria Control Centre, Lusaka (2011)
86. Duncombe, J., Clements, A., Hu, W., Weinstein, P., Ritchie, S., Espino, F. E.: Geographical information systems for dengue surveillance. Am. J. Trop. Med. Hyg. **86**(5), 735–755 (2008)
87. Kittayapong, P., Yoksan, S., Chansang, U., Chansang, C., Bhumiratana, A.: Suppression of dengue transmission by application of integrated vector control strategies at sero-positive GIS-based foci. Am. J. Trop. Med. Hyg. **78**(1), 70–76 (2008)
88. Cline, B.L.: Invited editorial for the inaugural issue of Geospatial Health. Geospat. Health **1**, 3–5 (2006)
89. ESRI: New Zealand Animal Health Board. ESRI, Redlands (2009)
90. ESRI: Safeguarding animal health more effectively. ArcUser Online (2005). http://www.esri.com/news/arcuser/1005/odaff.html. ESRI, Redlands
91. Wong, Y.I., Hillier, A.: Evaluating a community-based homelessness prevention program: A geographic information system approach. Admin. Soc. Work **25**, 21–45 (2001)
92. NASW Case Management Standards Workgroup: NASW Standards for Social Work Case Management. National Association of Social Workers, Washington (1992)
93. ESRI: There's no place like close to home, Illinois takes a geographic approach to foster care. HealthyGIS (2007). https://www.esri.com/news/arcnews/summer07articles/theres-no-place.html. ESRI, Redlands
94. MEASURE, WHO, Health Facility Assessment Technical Working Group: The Signature Domain and Geographic Coordinates:

A Standardized Approach for Uniquely Identifying a Health Facility. Univ. North Carolina, Chapel Hill (2007)
95. Kho, A., Johnston, K., Wilson, J., Wilson, S.J.: Implementing an animated geographic information system to investigate factors associated with nosocomial infections: A novel approach. Am. J. Infect. Control. **34**(9), 578–582 (2006)
96. PenBay Solutions: Indoor Building GIS Case Study for NYC Health. PenBay Solutions, Brunswick (2010)
97. Last, J. (ed.): A Dictionary of Epidemiology. Oxford Univ. Press, Oxford (2001)
98. Bell, S.B., Hoskins, R.E., Pickle, L.W., Wartenburg, D.: Current practices in spatial analysis of cancer data: Mapping health statistics to inform policymakers and the public. Int. J. Health Geogr. **5**, 49 (2006)
99. WHO: Global Observatory for eHealth (GOe) (2005). https://www.who.int/observatories/global-observatory-for-ehealth. WHO, Geneva
100. ESRI: HL7 and spatial interoperability standards for public health and healthcare delivery (2011). http://www.esri.com/library/whitepapers/pdfs/hl7-spatial-interoperability.pdf. ESRI, Redlands
101. ESRI: Interview with Ed Hammond (Chair of HL7). HealthyGIS (2009). http://www.esri.com/library/newsletters/healthygis/healthygis-winter2009.pdf
102. ESRI: Interview with K. Pemble, Wisconsin health information exchange. HealthyGIS (2010). http://www.esri.com/library/newsletters/healthygis/healthygis-winter2009-2010.pdf
103. Direct Relief International's Haiti Aid Distribution web application: Example for reporting tools. http://www.directrelief.org/Flash/HaitiShipments/Index.html
104. WHO: Indoor residual spraying: An operational manual for indoor residual spraying (IRS) for malaria transmission control and elimination. 2nd ed., Geneva. http://apps.who.int/iris/bitstream/handle/10665/177242/9789241508940_eng.pdf?sequence=1
105. Health Metrics Network: Framework and Standards for the Development of Country Health Information Systems. WHO, Geneva (2006)
106. Blatt, A.J.: Health, Science, and Place. Springer, Cham (2015)
107. Dent, D.M.: Cartographies of Disease. Maps, Mapping, and Medicine. ESRI, Redlands (2005)
108. Montgomery, C.W.: Environmental Geology, 11th edn. McGraw-Hill, Columbus (2019). Chap. 19
109. Palmer, G.T.: The diagnosis of the sick city. Am. J. Public Health **3**(7), 648–654 (1913)
110. Brownlee, S.: From pariah to pioneer. Dartmouth Medical School (2013). https://dartmed.dartmouth.edu/fall13/html/from_pariah_to_pioneer/
111. Davenhall, B.: Can geography make you healthy? TEDMED (2009). https://www.ted.com/speakers/bill_davenhall
112. DiChristina, M.: Why your doctor should know were you have lived. Scientific American (2009). https://blogs.scientificamerican.com/observations/why-your-doctor-should-know-where-you-have-lived/
113. Kornfield, A.B.E.: The missing piece in understanding our health. Huffington Post (2009). https://www.huffpost.com/entry/the-missing-piece-in-unde_b_336329
114. Terry, K.: Geomedicine new frontier in medical informatics. Information Week (2012). https://www.informationweek.com/mobile/geo-medicine-new-frontier-in-medical-informatics/d/d-id/1104732
115. Franklin-Wallace, O.: The geodoc will see you now. WIRED.UK (2014). https://www.wired.co.uk/article/geodoc
116. Davenhall, W.: Geomedicine: geography and personal health (2012). https://www.esri.com/library/ebooks/geomedicine.pdf. ESRI eBook, Redlands
117. Environmental Protection Agency: Toxic release inventory (TRI) (2019). www.epa.gov\tri
118. Boulos, M., DeBlond, J.: On the road to personalized and precision geomedicine. Int. J. Health Geogr. **15**, 5 (2016)
119. Duke Center for Health Informatics: Geomedicine summit (2014). https://dukeinformatics.org/conferences/2014-conferences/duke-geomedicine-summit-2014/
120. Kothari, C.R.: Research methodology: methods and techniques (2007). https://research-methodology.net/research-methods/. New Age, New Delhi
121. Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F., Billa, L.: Spatial analysis and GIS in the study of COVID-19. A review. Sci. Total Environ., **739**, 140033 (2020). https://doi.org/10.1016/j.scitotenv.2020.140033
122. Johns Hopkins University: Coronavirus Resource Center (2020–2021). https://coronavirus.jhu.edu/map.html. Johns Hopkins University, Baltimore, MD
123. CDC: COVID-19 vaccination equity (2021). https://covid.cdc.gov/covid-data-tracker/#vaccination-equity. CDC, Atlanta, GA

William F. Davenhall William Davenhall received his MA from the University of Kentucky, Department of Sociology & the Chandler Medical Center in 1974. He worked at Spindletop Research, Lexington; the Kentucky Hospital Association, the Methodist Evangelical Hospital; Human Services Coordination Alliance (HSCA), based in Jefferson County, Kentucky in the early 1970's; First Health Research, Louisville; Esri, Redlands, CA. He is retired and focuses on research projects relevant to the application of geomedicine in public and private healthcare settings.

Christopher Kinabrew Mr Kinabrew received an MPH/MSW from Tulane University's School of Public Health and Tropical Medicine and School of Social Work (New Orleans, LA). He was Esri's Public Health Specialist from 2005–2010. Currently, he is VP of Programs for the National Network of Public Health Institutes and Principal Investigator for cooperative agreements with the Centers for Disease Control and Prevention.

Open Source – GIS

30

Peter Löwe, Álvaro Anguix Alfaro, Andrea Antonello, Peter Baumann, Mario Carrera, Kim Durante, Marco Hugentobler, Steve Lime, Helena Mitasova, Dietmar Müller, Markus Neteler, Jack Reed, Christian Strobl, and Paul Wessel

Contents

P. Löwe (✉)
IT and eScience Department, WZB Berlin Social Science Center
Berlin, Germany
e-mail: peter.loewe@wzb.eu

Á. Anguix Alfaro · M. Carrera
gvSIG Association
Valencia, Spain
e-mail: aanguix@gvsig.com

M. Carrera
e-mail: mcarrera@gvsig.com

A. Antonello
HydroloGIS S.r.l.
Bolzano, Italy
e-mail: andrea.antonello@hydrologis.com

P. Baumann
Jacobs University
Bremen, Germany
e-mail: baumann@rasdaman.com

K. Durante
Science and Engineering Libraries, Stanford University
Stanford, CA, USA
e-mail: kdurante@stanford.edu

M. Hugentobler
Sourcepole
Zürich, Switzerland
e-mail: marco.hugentobler@sourcepole.ch

S. Lime
MnIT Services, State of Minnesota
St. Paul, MN, USA
e-mail: steve.lime@state.mn.us

H. Mitasova
Dept. of Marine, Earth and Atmospheric Sciences, North Carolina State University
Raleigh, NC, USA
e-mail: hmitaso@ncsu.edu

D. Müller
School of Geosciences, The University of Sydney
Sydney, NSW, Australia
e-mail: dietmar.muller@sydney.edu.au

M. Neteler
mundialis GmbH & Co. KG
Bonn, Germany
e-mail: neteler@mundialis.de

J. Reed
Digital Library Systems and Services, Stanford University
Stanford, CA, USA
e-mail: phillipjreed@gmail.com

C. Strobl
IT Strategy, eGovernment and Applications, Bavarian State Ministry of the Environment and Consumer Protection
Munich, Germany
e-mail: christian.strobl@stmuv.bayern.de

P. Wessel
Dept. of Earth Sciences, University of Hawaii at Mānoa
Honolulu, HI, USA
e-mail: pwessel@hawaii.edu

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_30

Abstract

This chapter provides an overview over the current spectrum of free and open source licensed geospatial software tools and communities.

The number of available open source geospatial tools continues to grow and diversify, while new fields of application emerge. This chapter presents a selection of well-established free and open source software (FOSS) tools as reference cases, including software libraries, desktop applications, web mapping, databases, and use with portable devices, mostly based on the portfolio of FOSS tools, which is curated and maintained by the metaproject OSGeoLive, which is also discussed in this chapter.

FOSS software options are available for almost any kind and magnitude of geospatial task, for some niche tasks, they can even be the only options. Geospatial FOSS are available for portable devices, desktop machines, high-performance cluster environments, as Web-based applications or Software-as-a-Services (SaaS) in cloud environments. While a range of different open source licenses exists and are applied to geospatial software, they all adhere to the same key principles of openness and freedom for the user, most notably the rights to run the program, access the source code for study and improvements, and to redistribute copies of the program and its modifications.

This allows for transparency of the computations, as the source code implementation of all algorithms can be verified and possibly improved when necessary. Also public access to the code base ensures sustainable long term availability of the software tool to the users.

FOSS geospatial software has proven to be just as reliable and suitable for professional use as closed-source software alternatives. As a service to the readers, the reference section of this chapter includes scientific video recordings which provide additional content regarding software capabilities, historical notes and background information: Actinia [1], GeoBlackLight [2], GeoPaparazzi [3, 4], GeoPython [5–7], GeoServer [8, 9], GeoTools [9], GMT [10], GRASS GIS [1, 11–13], gvSIG [3, 14–16], JTS [17], Leaflet [18], MapServer [19, 20], NASA World Wind [16], OSGeoLive [21], PostGIS [22], PROJ [23, 24], QGIS [25], rasdaman [26]. In addition, scientific videos about the OSGeo Foundation [27] and the annual Sol Katz Awards or Geospatial Free and Open Source Software [28, 29] are included.

Keywords

Open source · GIS · FOSS · free software · OSGeo · geoinformatics

30.1 History of FOSS Geospatial Tools

The evolution of todays' ecotope of FOSS geospatial tools and communities is more than just a sequence of independent software tools, providing a growing range of functionalities to their users, accelerated by the decline of computer hardware prices and increasing processing power. While these factors are relevant, others must also be acknowledged. From its beginnings, geospatial FOSS tools were developed jointly by groups of stakeholders, motivated by their occupational duties, self-education, research or commercial motivation: Collaboration, mutual sharing of resources and social codes of conduct have become key aspects to all successful FOSS projects. Like Wikipedia, they are digital commons, where information and knowledge resources are collectively created, owned, and shared among a community, also being freely available to third parties.

The origins of this development date back to 1978, when the implementation of the first public domain vector-based and command line driven geographic information system (GIS) Map Overlay and Statistical System (MOSS) was initiated by the US Fish and Wildlife Service (US FWS) [30, 31].

Active development of MOSS continued up until 1987, following the technological evolution from mainframe computing to microcomputers, and finally personal computers (PC MOSS).

The codebase and documentation of PC MOSS is still available as both an Open Access software publication [32] and a software repository [33, 34]. PC MOSS can still be used today through PC emulation software. MOSS is recognized as an open geospatial heritage project by the Open Source Geospatial Foundation (OSGeo) [35].

The open source project to emerge next was GRASS GIS (Geographic Resource and Analysis Support System) in 1982 [36].

GRASS GIS is currently the geospatial FOSS project with the longest continuing development which still continues today [36]. It served as a best practice example for later FOSS geospatial projects and communities including OSGeo. Development was launched in 1982 by the U.S. Army Corps of Engineers' Construction Engineering Research Laboratory (USA-CERL). GRASS GIS, an originally Unix-based raster GIS was put in the public domain, similar to MOSS. A user base developed among US federal agencies, as well as in academia in the US and internationally.

Organizational bodies were created, including the Open GRASS Foundation (OGF). For the users, a printed GRASS community magazine was published, annual user meetings were held from 1985, and a video commercial was produced [37].

Predating the public availability of the WWW in 1993, GRASS GIS was among the first open source development projects using the Internet for communication, with more than 6000 subscribers to the GRASS mailing lists for developers and users in the early 1990s. In 1997, project management and the code repository were transferred from USA-CERL to academia, where the software was ported to Linux.

Beginning in 1994, OGF shifted its focus and became the Open GIS Consortium (OGC), renamed as Open Geospatial Consortium (OGC) in 2004, a consortium to foster open standards for geospatial interoperability, a key factor for the client-server services like Web Map Service (WMS), Web Feature Service (WFS), and Web Coverage Service (WCS) according to ISO 19128: "Server", to support ubiquitous, standards-based geoprocessing.

Today, OGC is an international voluntary consensus standards organization, encouraging development and implementation of open standards for geospatial content and services, sensor web, GIS data processing, and data sharing.

Due the rise of personal computers and the growth of academic Internet access, by the end of the 1990s, the GRASS GIS code base on the Internet was discovered by a new group of stakeholders, students, which required alternatives to the restricted access to GIS software at universities at that time. They revived the code base, communicating via the established mailing lists, which were now managed by volunteers. After a gap of 6 years, GRASS conferences were held again after 2000.

The primary code repository was maintained from 1998 at the University of Hannover, Germany, and from 2001 to 2006 at ITC-irst in Trento, Italy and from then on by the OSGeo Foundation, US.

The license of the GRASS source code was changed in 1999 from the public domain to the GNU General Public License (GPL), ensuring the defined freedom of open and free software. This sparked a renaissance of development activities and allowed the GRASS Development Team to grow into a multi-national team [38]. Also, companies started to provide professional services based on GRASS GIS.

The improving IT best practices for code repository management were taken up by the project community early on, by introducing the code versioning system CVS (Concurrent Versions System) in 1999, moving on to SVN (subversion) in 2007, followed by GitHub, starting in 2019. Since 2021, the GRASS GIS repository is long term preserved by the Open Access repository Zenodo, and scientifically citable by DOI [38].

In 2006, GRASS GIS was among the founding projects of the OSGeo Foundation, which will be described later.

In the meantime, further FOSS projects were started. In 1983, the development of the PROJ library started, based on earlier efforts [39], which was evolved from earlier FORTRAN code [40].

The map generator project Generic Mapping Tools (GMT) was launched in 1988. Both the MapServer and rasdaman projects were started in 1995.

Since the mid-1990s, the pace of the appearance of new projects quickened, and it has continued to accelerate, with at least one new project annually. The following summary of the development is incomplete and focuses on the geospatial software projects presented later in this chapter.

In 2000, the Geospatial Data Abstraction Library (GDAL), a cross-platform geospatial library, was started, which would in time become the most used geospatial data access library by both FOSS and proprietary geospatial tools.

PostGIS, a spatial extension for the PostgreSQL database system, and GeoServer were begun in 2001, the desktop geographic information systems QGIS in 2002 and gvSIG 2003 in 2003, and GPlates in 2006. Since 2008, the OSGeoLive project has been providing annual integrated ready-to-run collections of FOSS geospatial tools and open data, continuously updating and renewing its software portfolio for new tools, such as the Leaflet library for web mapping applications, which started in 2011.

Due to the freedom provided by FOSS licenses, FOSS geospatial projects soon started to collaborate and share resources by mutual embedding of functionalities and underlying code. Prime examples are specialized core FOSS software libraries like GDAL/OGR or PROJ, which have become integrated in other FOSS projects like QGIS and MapServer and even closed source proprietary software.

Open source geospatial applications are available for the use in cluster-based high-performance computation, client–server scenarios, desktops, tablets, and smartphones, and single-board microcontrollers like Adruino.

This range of FOSS geospatial tools enables the processing of a plethora of input data formats, including vector and raster data, but also voxel-based volumes, multidimensional complex data cubes, temporal data and data streams from SWE-enabled environmental sensors.

On the organizational level, the communication and coordination among the geospatial FOSS projects improved significantly over time [41].

By the early 2000s, apart from OGC, which advances open standards applicable for all GIS software, the stakeholder groups of the individual FOSS geospatial projects primarily communicated among themselves, with only limited cross-project exchange.

As a next stage, cross-project communication was established on the national scale, which led to the introduction of national volunteer organizations like the FOSSGIS e. V. association in Germany and the Associazione Italiana per l'Informazione Geografica Libera in Italy.

International multiproject communication picked up in 2004 with the first international conferences on FOSS GIS in Canada (Open Source GIS conference) and Thailand (Free/Libre and Open Source Software for Geoinformatics: GIS-GRASS Users Conference), where the acronym FOSS4G was coined [42], which since 2006 has become the brand for national and global FOSS conferences.

In 2005, an initial proposal for a MapServer-focused foundation was opened up towards other geospatial FOSS projects, which led to the establishment of the international Open Source Geospatial Foundation (OSGeo) in 2006.

OSGeo is a not-for-profit organization devoted to an open philosophy and participatory community-driven development to provide financial, organizational, and legal support to the broader Free and Open-Source Software geospatial community [27].

OSGeo projects offer freely available tools and technologies under an open source license. The OSGeo web portal promotes the work of teams and organizations worldwide, while volunteers are organized in over 20 local national or regional chapters and initiatives, reaching out to the GIS industry, education, and academia.

OSGeo draws governance inspiration from several aspects of the Apache Foundation, including a membership composed of individuals drawn from foundation projects. Individuals are selected for membership status based on their active contribution to foundation projects and governance. The foundation coordinates the efforts of its member projects and establishes quality standards through assessments and mentorship through the OSGeo incubation process. From 2006 until 2022, 21 geospatial FOSS projects, including content management systems, desktop applications, geospatial libraries, metadata catalogues, spatial databases, and web mapping, have graduated from the OSGeo incubation process [43]. The incubation process continues to evolve and embraces technological advances and emerging best practices, like software citation [44]. This provides guidance for new users and developers, ensuring that these OSGeo projects are dependable, reliable, and well governed concerning their software, documentation, and community standards.

By 2020, 454 charter members had been elected, comprising software developers, activists, and advocates [45] and over 36 000 persons subscribing to the OSGeo mailing lists [46].

Since 2006, OSGeo has been hosting annual Free and Open Source Software for Geospatial (FOSS4G) Conferences, both as global events and also as independent national events. In 2021, over 20 regional conferences and meetings were held [47].

The beginnings of the FOSS geospatial ecotope are annually remembered at the International FOSS4G Conference by the awarding of the Sol Katz Award for Geospatial Free and Open Source Software [28, 29, 48] for individuals who have advanced the open source ideals in the geospatial realm. The award honors the late Sol Katz, an early pioneer of FOSS Geospatial, who was involved in the development of the MOSS project.

30.2 Free and Open Source Licenses

Software, like any other original creative work, is generally copyright protected, unless it is specifically made available in the public domain, where no exclusive intellectual property rights apply. If a software is to be used by persons other than the copyright holder, a license must be granted to the user.

Before the 1980s, due to the significant prices for computer hardware, software was distributed both as source code and free of cost since it was considered part of the computer system but not as an individual product. In that time, the purchasers of computers, such as universities or research institutes, altered the software by adding new functionalities by extending the source code. This is similar to what is done today with open source software, yet the concept had not been defined.

The concept of software licenses dates back to the year 1980, when the term computer software was added to US copyright legislation, which was adopted in the following years by many other nations. Currently, software falls into two broad categories: proprietary software and free/open software. The difference between both types are the applicable licensing terms.

A software license is a legal instrument that governs how the software is to be used and distributed. Software licenses can be either proprietary or open. Proprietary software is copyrighted and is licensed in such a way that use, distribution, and modification of the code are limited by the publisher, vendor, or developer. Proprietary software remains the property of its owner/creator and is used by end-users/organizations under predefined conditions. The original source code for the proprietary software is not available. Because of this it is also referred to as closed source software. The refusal to provide access to the source code is a measure to protect the intellectual property invested in software development. Many proprietary software applications are also commercial, meaning that they are produced for sale, and users have to pay for a license. However, proprietary software can also be free of charge: Shareware is an example of proprietary software that is provided to users at no cost, yet under certain conditions (e.g., limited functionality, expiration after a trial period.). The fact that software is provided free of charge does not mean it is not proprietary.

In the 1980s, the emergence of a commercial software industry based on proprietary closed source software products led to a countermovement to continue the earlier established practice to study, share, and modify software by the introduction of free and open software licenses. The term free is used in the sense of "free speech" (libre), not as in "free of charge" (gratis). The free software definition [49] defines free software as software that ensures that end users have freedom in using, studying, sharing, and modifying that software.

The modern definition of free software comprises the following four types of freedom [50]:

- The freedom to run the program, for any purpose (freedom 0).
- The freedom to study how the program works and change it for new tasks (freedom 1). Access to the source code is a precondition for this.
- The freedom to redistribute copies so others can benefit, too (freedom 2).
- The freedom to distribute copies of modified versions to others (freedom 3). This gives the whole community a chance to benefit from the changes made.

Freedom 1 and freedom 3 require access to the source code.

Free software is often developed by collaborative teams, where many users and developers contribute to ongoing improvements of a software project. To recognize this feature of community-involvement, the new term open source was defined in 1998 by the Open Source Initiative (OSI), based on the Debian Free Software Guidelines [51], complementing the concept of free software.

For software that is both free and open, the terms free and open source software (FOSS) and free, libre, and open source software (FLOSS) were coined.

Cases exist where copyright owners provide dual licensing to allow user communities to use the software either under a FOSS or proprietary license. Over the past decades, a growing number of both open and free software licenses have been defined. They can be categorized based on their permissivity, granting, or retaining rights by the copyright holders.

Permissive licenses, also referred as non-copyleft licenses, impose minimal restrictions on the use and distribution of covered software. Some do not require the provision of the source code, which is a deviation from the freedoms 1 and 3 for free software. A benefit of permissive licenses is that they enable compatibility with software published under other permissive licenses. The most widely used permissive licenses are the Apache 2.0, BSD, and MIT/X11 licenses.

The Apache 2.0 License, published by the Apache Software foundation in 2004, precedes the earlier Apache 1.1 License from 2000 [52]. Apache 2.0 is compatible with the GPL 3.0 license but not earlier versions of GPL. It requires preservation of both copyright notice and disclaimer. It is used by the Geoblacklight project. The acronym BSD is derived from Berkeley Software Distribution, a Unix-like operating system. A variety of BSD license types has evolved from the original license definition from 1988 [53]. All BSD licenses require that all software retain the BSD license notice when redistributed as source code or display the notice when distributed in binary format. Source code distribution is optional. Of the software tools in this chapter, Leaflet, Geopython, and JTS have BSD-based licenses.

The MIT License, also referred to as X-License or Xl I-License, was published by the Massachusetts Institute of Technology (MIT) in 1988 [54]. MIT-licensed code is compatible with copyleft licenses and can be integrated into GPL-licensed software. In this chapter, the MapServer, PROJ.4, and GDAL/OGR projects use MIT-compatible licenses.

Less permissive licenses are referred to as copyleft licenses. They are more protective than permissive licenses by maintaining the identical distribution rights for derivative works. They enforce the publication of the source code whenever the software is being redistributed. This ensures that future versions of the software will remain free and publicly available.

The majority of the software projects described in this chapter apply varieties of the General Public License (GPL), a family of licenses created by the Free Software Foundation (FSF), first published in 1989 [55]. The current version is GPL 3.0, which was released in 2007 [56].

GPL-derived licenses cover weak and strong versions of copyleft. The GPL3 Affero (AGPL) license provides strong copyleft, as the source code has to be published even when the software is deployed as software-as-a-service. Weak copyleft is provided by the GNU Lesser General Public License (LGPL), which allows LGPL-licensed code to be used as shared software libraries for permissively licensed or proprietary software.

QGIS, GRASS GIS, PostGIS, Geopaparazzi, GeoServer, and GPlates use GPL-type licenses. gvSIG deploys the AGPL license, GMT and Geotools use LGPL licenses, while other projects use GPL licenses. rasdaman uses LGPL for the client parts – thereby allowing inclusion in any software – while the core server engine (running as a separate process) is available under a GPL license.

In research and education, FOSS has become one requirement for the paradigm shift towards Open Science, consisting of Open Access to scientific articles, Open Data, and Open Source:

Access to software source code and the sharing of the programs creates trust in both research findings and software tools by verification and replication of scientific results. It also reduces redundant development efforts, speeds up development efforts, and fosters innovation.

30.3 Desktop GIS

The term desktop GIS stands for general purpose GIS locally installed on a desktop or laptop computer, which is used for viewing, editing, and analysis. Desktop GIS can be connected with other systems, such as Data Stores or Web mapping Services.

30.3.1 QGIS

QGIS is an open source GIS, available under the terms of the GNU General Public License, version 3 [57]. The software was originally called Quantum GIS and changed the name to QGIS with the release of version 2.0 in 2013. QGIS is written in C++ with language bindings for Python. The "Q" in the name is because QGIS uses the Qt libraries [58]. The source code of QGIS is hosted on the GitHub platform [59], and binaries are available for most platforms, including different Linux distributions, Microsoft Windows and Mac OS. It is also available as an Open Access publication [60].

The QGIS project was founded by Gary Sherman in 2002 with the aim of providing a user-friendly GIS providing common functions and features to access data stored in PostGIS databases. Since 2007, the QGIS project has been a member of the OSGeo foundation. Version 1.0 was released in 2009, version 2.0 in 2013, and version 3.0 in 2018 [25]. A new version is released every 4 months, and one release per year is a long-term release (LTR), therefore receiving bugfixes for 1 year.

Supported GIS Formats

QGIS uses the GDAL/OGR library to read vector and raster data, and therefore supports hundreds of different GIS formats. For some data sources like PostGIS, Oracle, GPX, Text, or WMS, QGIS has developed data drivers and does not use the GDAL library.

Functionality

Viewer

Displaying geodata is one of the strong points of QGIS. Vector and raster layers can be symbolized in a variety of ways. Besides the cartographic functions available in most GIS software, QGIS also supports advanced visualizations, like clustered rendering of points, live layer effects, or a combination of layers using different blend modes (Fig. 30.1).

3-D visualization was provided by plugins in QGIS2. Since QGIS3, a 3D viewer is integrated in QGIS by default.

Editor

Vector data can be edited in the attribute table or in the map view. The functions available in the map view include adding new objects, editing vertices of objects, simplifying objects, offset lines, and split features, combining features, reverting lines. Undo/redo is available for all editing functions.

Print Composer

In the print composer module, print layouts can be creating by placing items (e.g., maps, legends, scalebars, text, and images) on a canvas. The creation of reports and serial prints are also supported.

Processing

The processing framework contains a large number of analysis functions. These functions can be combined in a graphical modeler to create complex processes.

Plugins

New functions can dynamically be added to QGIS with plugins written in C++ or Python. Python plugins can be managed and distributed with plugin repositories. The official repository of the QGIS project [61] is publicly accessible and contains over 900 plugins.

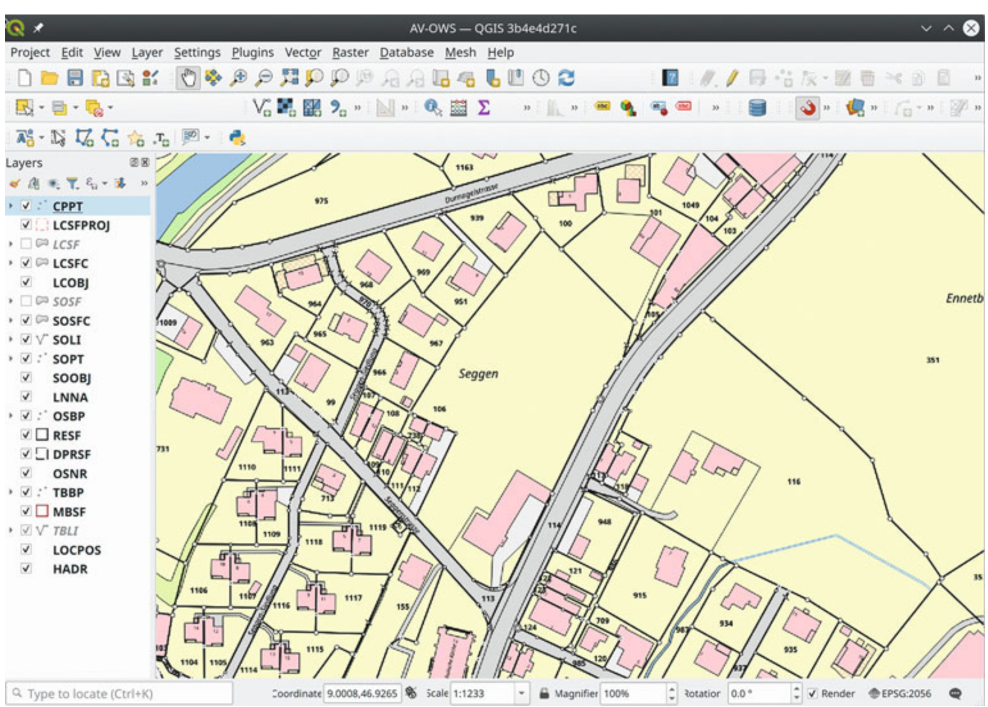


Fig. 30.1 The QGIS user interface including the layers panel (*left*). The map display shows a thematic mapping of buildings, lots and roads in Linthal, Switzerland. QGIS has been translated into over 85 languages

Fig. 30.2 Examples of geospatial modeling and analysis in GRASS GIS with 2-D/3-D raster data, vector data, and tangible interaction. Locations: (*from left to right*) Yakima, Washington, USA; Jockey's Rifge sand dune, North Carolina, USA; Rayleigh, Wake County, North Carolina, USA

Server
QGIS Server provides an Open Geospatial Consortium (OGC) WMS, WFS, and WCS server on top of the QGIS libraries. A nice feature of QGIS Server is that projects' layer symbolization can be configured in a graphical way by using QGIS desktop. QGIS Server provides some extensions to the WMS standard, like a GetPrint request for web-based printing and sketching functionality.

The development of QGIS server started in 2006 at the Institute of Cartography, ETH Zurich. In 2010, the source code was merged into the official QGIS source code repository.

30.3.2 GRASS GIS

GRASS GIS (https://grass.osgeo.org) [36] is a geospatial software suite originally developed in 1980–1990s by the US Army Corps of Engineers Construction Engineering Research Laboratories (USA-CERL) in Champaign, Illinois [36, 62] to support land management at US military installations [37]. After USA-CERL ceased its official support of GRASS in 1996, a new development team was formed. This new team adopted the GNU GPL license in 1999. In 2006, GRASS GIS became one of the founding projects of the newly established OSGeo Foundation. According to *Westervelt* [62], several concepts contributed to the long-term success of GRASS: novel GIS analysis approaches were needed, and only an open system provided the opportunity for the efficient development of new software tools; and although the core system was maintained by a core group, anyone interested could improve existing modules or develop new ones [11].

Throughout its 35+ years, the GRASS GIS software development has been driven by scientists, engineers, graduate students, and practitioners with the aim to provide tools for geospatial data management and analysis, image processing, spatial modeling, and visualization (Fig. 30.2) [12]. GRASS GIS evolved into a general-purpose, portable GIS software running on all common operating systems, from desktop to high-performance computing (HPC) and cloud installations. Following the Unix philosophy that each module does a specific task, GRASS GIS is a system of over 400 modules where each module is autonomous, including memory management and error handling. The core libraries and most modules are written in POSIX-conforming ANSI C. Some functions are written in C++ and increasingly in Python [63]. GRASS GIS offers flexibility through its interfaces, by supporting command-line and graphical user interfaces, along with a Python API and an innovative tangible interface [64].

The add-ons repository broadens the contributors community and serves as a testing environment, since add-ons that are highly used are eventually moved to the core repository. The evolution of the GRASS codebase within the repository has itself become an object of analysis and research [13]. Furthermore, a "sandbox" repository is used for sharing the development of highly experimental or nonstandard tools. Collaboration with other open source projects, such as QGIS, R, or GDAL ensures interoperability. GRASS GIS has always served as a geospatial innovation platform with a history of unique contributions, from 3D dynamic visualization and global watershed analysis in the early 1990s to the more recent development of geomorphons [65], temporal framework [66], or tangible modeling [67]. Moreover, integrating scientific algorithms into GRASS GIS helps to preserve reproducibility of scientific results over time [68]. In addition, actinia [1, 69, 70], a REST API for scalable, distributed, high-performance processing of geographical data that uses mainly GRASS GIS along with GDAL/PROJ/SNAP, as well as the possibility to deploy own scripts wrapped as GRASS GIS add-ons, has been developed. Actinia consists of several components:

i) actinia-core (available at https://github.com/mundialis/actinia_core) for the processing of single and time series of satellite images, of raster and vector data.
ii) actinia-gdi with an interface to Geonetwork Open Source for the access to a metadata catalogue (available at https://github.com/mundialis/actinia-gdi).
iii) actinia plugins for domain-specific applications.

The cloud deployment of actinia is possible through docker (docker-compose and docker-swarm) as well as Helm charts in OpenShift, OpenStack servers, and Kubernetes clusters. actinia has been an OSGeo Community Project since 2019.

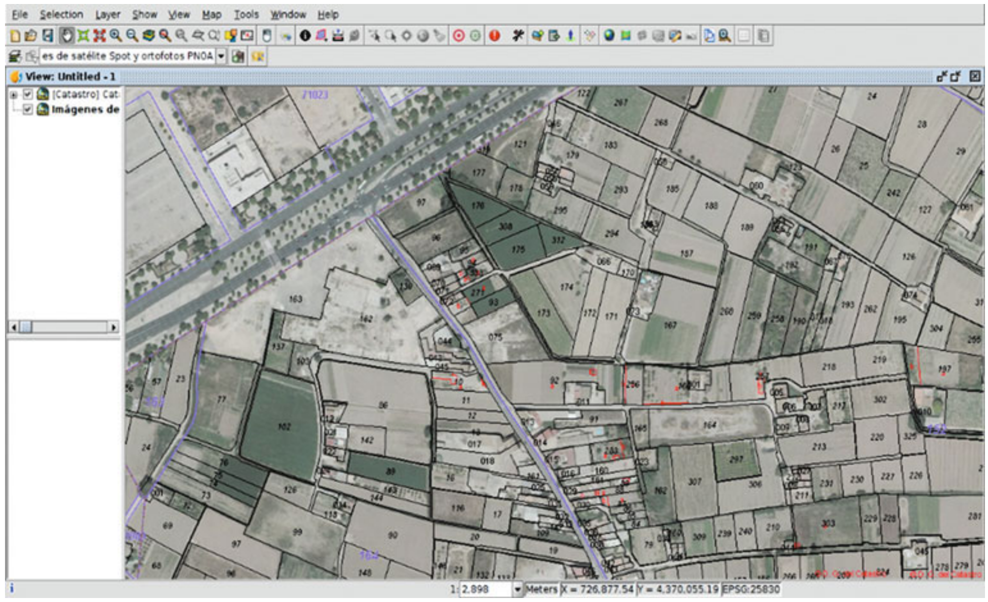


Fig. 30.3 View overlapping local and remote data from Cadastre WMS and another WMS about Spanish spatial data infrastructure. Location: Alcàsser, Spain

GRASS GIS applications cover many scientific and engineering fields, including terrain mapping, analysis and modeling in archeology, agriculture, and geosciences, analysis of land cover and its structure, simulation of landscape processes, renewable energy assessment, network analysis, and many others [71, 72].

30.3.3 gvSIG, Open Source Software for Geomatics

The gvSIG project [14, 73, 74] started with a desktop geographic information system, but currently that application is part of the **gvSIG Suite**, a whole catalogue of open source software solutions. The gvSIG Suite is composed of "horizontal" solutions and a wide range of sector products. The current products of the gvSIG Suite:
"Horizontal" products:

- gvSIG Desktop: geographic information system for editing, 3-D analysis, geoprocessing, maps, etc.
- gvSIG Online: integral platform for Spatial Data Infrastructure (SDI) implementation.
- gvSIG Mobile: mobile application for Android to collect field data.

Sector products:

- gvSIG Educa: gvSIG adapted to geography learning in preuniversity education.
- gvSIG Crime: geographic information system for criminology management.
- gvSIG Roads: platform to manage road inventory and conservation.

The project was born in Generalitat Valenciana (the regional government in Valencia, Spain) in 2002. Until 2009 it was the only organization that managed the project, but in 2010, the gvSIG Association was born to guarantee the sustainability of the project. The gvSIG Association is a nonprofit association composed of SMEs and nonbusiness organizations (universities, public administrations, technological institutes, etc.). The objective of the Association is that the "benefits" of its activity reverts for the sustainability of the project. gvSIG project graduated as an OSGeo project (the Open Source Geospatial Foundation) as of November 2015.

gvSIG Desktop. Desktop GIS Application

gvSIG Desktop [15] is considered a powerful GIS and SDI client, which supports most known data formats, raster and vector formats, like shapefile, dxf, dwg, dgn, and most of the geospatial databases like PostGIS, mySQL, Oracle, H2, Spatialite, and SQLite It provides the most common GIS tools like data loading, map navigation, and query map information like alphanumeric information, distance measurement, thematic cartography, legend editing using the most common legend types, labeling, feature selection, attribute tables with statistics, ordering, table relations, table linking, layout manager, geoprocessing tools, CAD, raster processing, etc. (Fig. 30.3).

gvSIG Desktop [75] is available in more than 30 languages, and all of them have been translated by the gvSIG community. The application is multiplatform, and there are different distributions for Linux, Windows, and Mac OS X, including portable and installable versions.

There is also a scripting module that allows customization of gvSIG and the development of new tools in an easy way due to the use of Python [76].

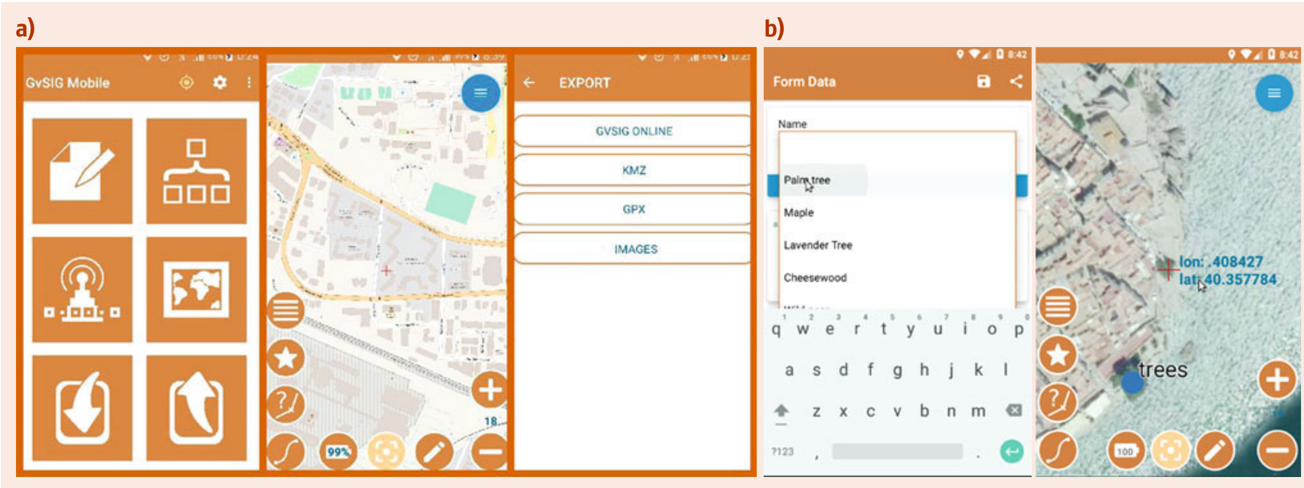


Fig. 30.4 **a** Different windows of the gvSIG Mobile application: main menu, viewer, and export menu. **b** Customized forms in the gvSIG Mobile application for data gathering. Tree mapping in Peníscola (Spain)

Its SDI client condition allows connecting, through the use of standards, to Open Geospatial Consortium (OGC) Services like OGC WMS (also ISO 19128) (raster and vector data returned as georeferenced map images), OGC WMTS (raster and vector data returned as tiled map images), OGC WFS (also ISO 19142 and ISO 19143) (advanced access to vector data), and OGC WCS (advanced access to raster information). Moreover, apart from OGC services, from gvSIG users can access other types of services like OpenStreetMap, Google Maps, and Bing Maps, and it is also possible to open Google Street View from a gvSIG project.

The gvSIG Desktop application is free of charge, under GNU/GPL3 license, and it is available to download from the gvSIG project website.

gvSIG Mobile

gvSIG Mobile is a GIS application to be used for field work. It is a powerful, easy-to-use, and interoperable solution for data collection on Android devices. It is a cornerstone of gvSIG Suite and is integrated directly with gvSIG Desktop and gvSIG Online. It is oriented to data field collection and is recommended for inventory projects, census, revisions, and inspections.

The new gvSIG Mobile, released at the end of 2017 and developed under GNU/GPL license, is based on the Geopaparazzi application [77], and was born with the aim of having a mobile GIS application for professionals. It has different tools that facilitate its integration with the rest of the gvSIG Suite. For example, it has a data importer and exporter from/to gvSIG Online, functionality that is already used by many of the organizations betting on implementing their spatial data infrastructures (SDI) with this platform.

Of course, gvSIG Mobile can be used independently of the rest of the gvSIG Suite components. At an individual level, it is a fantastic application for field data gathering. It includes a lot of functionalities, but it is very easy to use. Users can gather field data, edit existing data, and attach images, notes, or bookmarks to geolocated elements, etc. (Fig. 30.4).

Customized forms can be created in gvSIG Desktop to be exported to gvSIG Mobile. They can include pictures, sketches, drop-down fields, among others.

gvSIG Mobile is free of charge, and it is available to download from Google Play.

gvSIG Online

gvSIG Online is the open source software suite to implement spatial data infrastructures that the gvSIG Association has driven strongly in the world of geomatics. It is very useful for municipalities, regional governments, ministries, or private companies, in conclusion, for any entity who needs to publish their geographic information in a geoportal or manage the cartography internally between different departments.

It is also common to find that products in the SDI technology market have a lot of restrictions.

These limitations on information layers, the number of queries, or the number of users let the user find products where the management capabilities of their geographic information are limited.

All these determinants have been taken into account to define gvSIG Online to provide solutions for the problems. gvSIG Online allows creating a geoportal in less than 5 minutes. The platform includes a file manager where vector and raster files can be uploaded and published in an easy way (Fig. 30.5). gvSIG Online is based on open source software, without the costs of license or maintenance. It is developed under AGPL (Affero) license [78]. All information to implement gvSIG Online is available at the gvSIG project website.

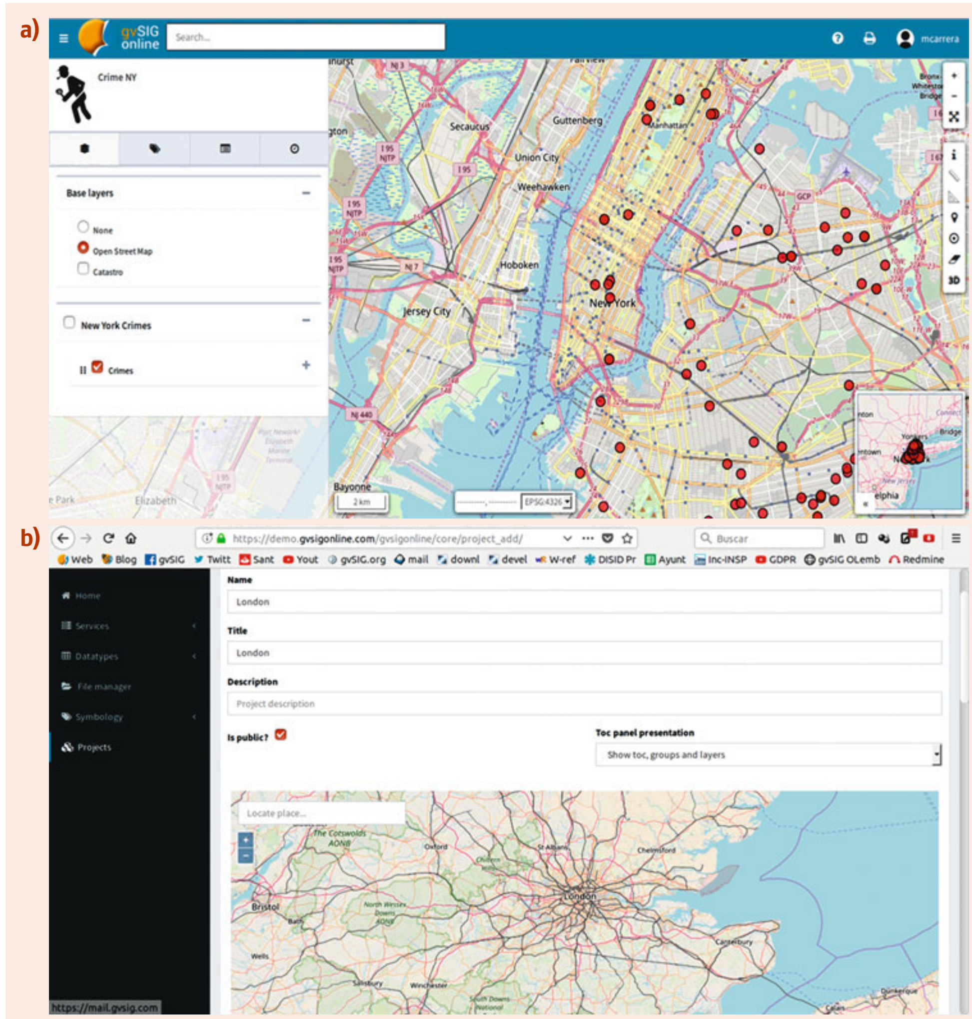


Fig. 30.5 **a** gvSIG Online: viewer, layers, and main tools. **b** gvSIG Online: creating a geoportal in a few minutes

30.4 Data Stores and Datacubes

Data stores and datacubes are geospatial repositories which focus on persistently storing and managing collections of spatial data on scales beyond the limited memory resources of desktop computers and the computing capabilities of single user GIS applications. Software applications for these tasks require database engines capable to manage very large and potentially heterogenous spatial data volumes. Datacubes are enabled by software providing web-based processing and querying capabilities for massive multidimensional arrays of gridded data.

30.4.1 PostGIS – an Open Source Spatial Database

PostGIS [79, 80] is a spatial database extension for the PostgreSQL object-relational database. Both PostGIS and PostgreSQL are open source software. PostGIS is released under the GNU General Public License [81] and PostgreSQL is released under the BSD license. The Open Geospatial Consortium OGC has certified PostGIS as a "Simple Features for SQL" compliant database [82]. Support for the ISO SQL/MM suite of spatial database functions is also implemented [83].

PostGIS adds geometry data types and spatial functions to the PostgreSQL database. The supported geometry data types are "Points", "LineStrings", "Polygons", "MultiPoints", "MultiLineStrings", "MultiPolygons", and "GeometryCollections" (Figs. 30.6 and 30.7). Spatial functions enable the analysis and processing of GIS objects. Examples are measurement functions like "Area", "Distance", "Length", and "Perimeter", and spatial operators (Fig. 30.8) like "Union", "Difference", "Symmetric Difference", and "Buffer". Topological relationships, like "Equals", "Disjoint", "Intersects", "Touches", "Crosses", "Within", "Contains", and "Overlaps" are processed by the Dimensionally Extended Nine-Intersection Model (DE-9IM). Since PostGIS 2.0, full raster support is added to the basic PostGIS distribution. Besides these there are multiple optional extensions, like the PostGIS Topology extension or third-party

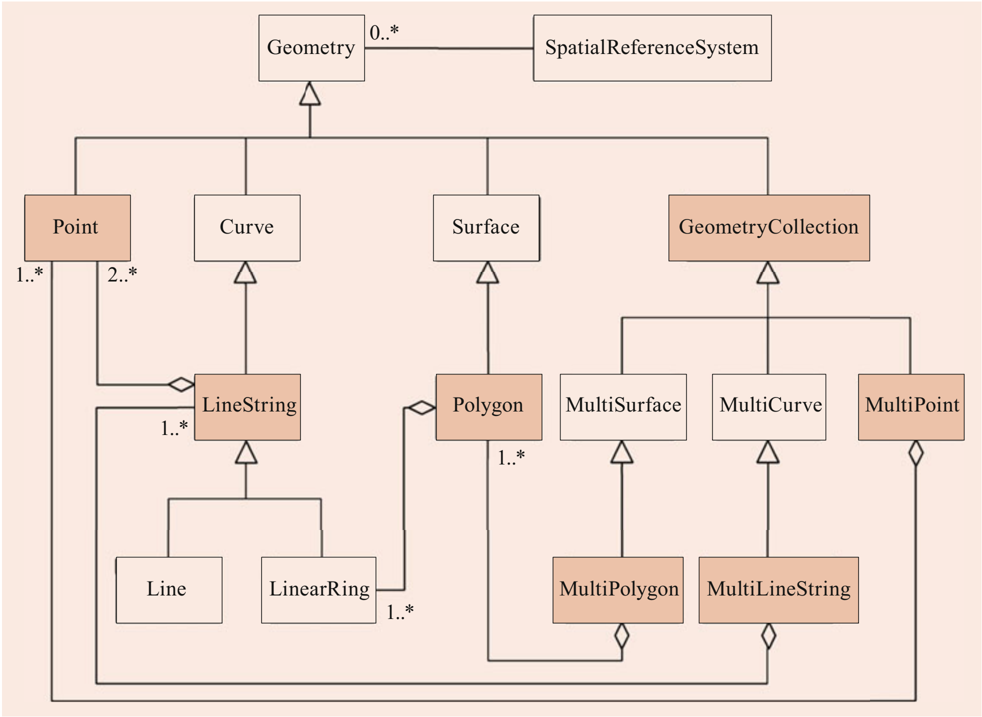


Fig. 30.6 Geometry Class Hierarchy of the *Simple Features for SQL* specification from the Open Geospatial Consortium. The Geometry Types supported by PostGIS are filled with a *darker color*

software like pgRouting [22], which extends the PostGIS/-PostgreSQL geospatial database to provide geospatial routing functionality.

In general, the functionality of PostGIS is comparable to Esri ArcSDE, Oracle Spatial, and DB II spatial extender. Like other spatial databases [84], PostGIS combines the advantages of classical GIS software, mainly the possibility of spatial analysis [85], with the advantages of database management systems (DBMS) like indexing, transactions, and concurrency [86, 87].

The implementation of the Open Geospatial Consortium (OGC) "Simple Features for SQL" [82] provides powerful features for managing, retrieving, and analyzing geospatial data. The spatial domain introduces a set of functions to the SQL language. A typical spatial query is provided in Fig. 30.9.

Spatial data infrastructures (SDI) facilitate the access to geospatial information using a minimum set of standard practices, protocols, and specifications [88]. Every SDI requires a spatial database server, and PostGIS represents

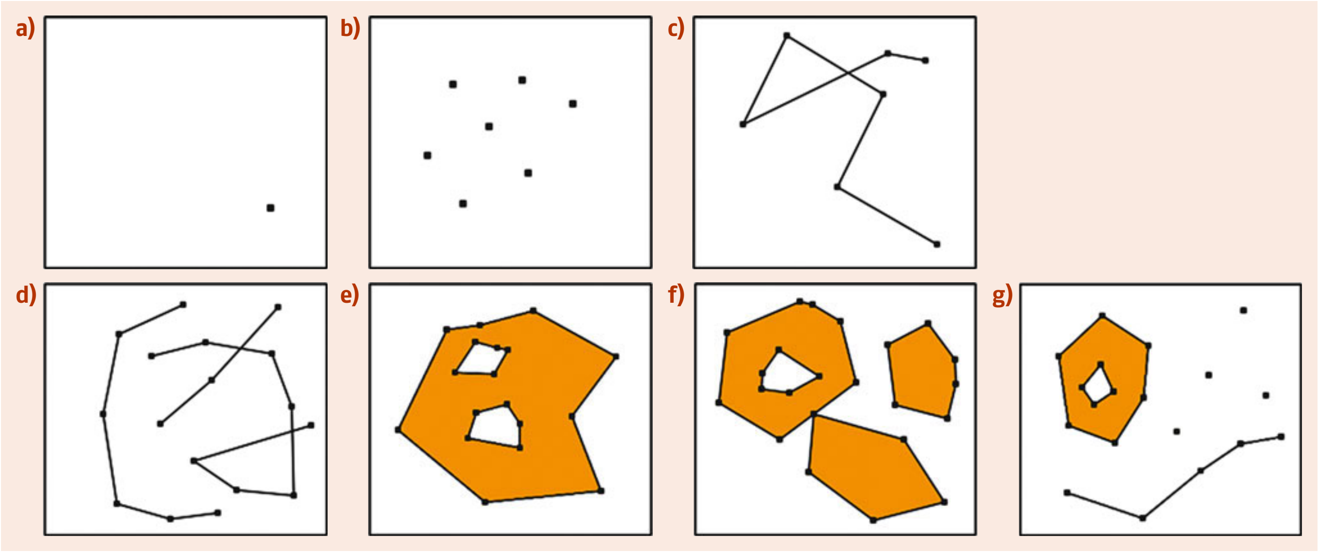


Fig. 30.7 Geometry Types supported by PostGIS: Point (**a**), MultiPoint (**b**), LineString (**c**), MultiLineString (**d**), Polygon (**e**), MultiPolygon (**f**), and GeometryCollection (**g**)

30

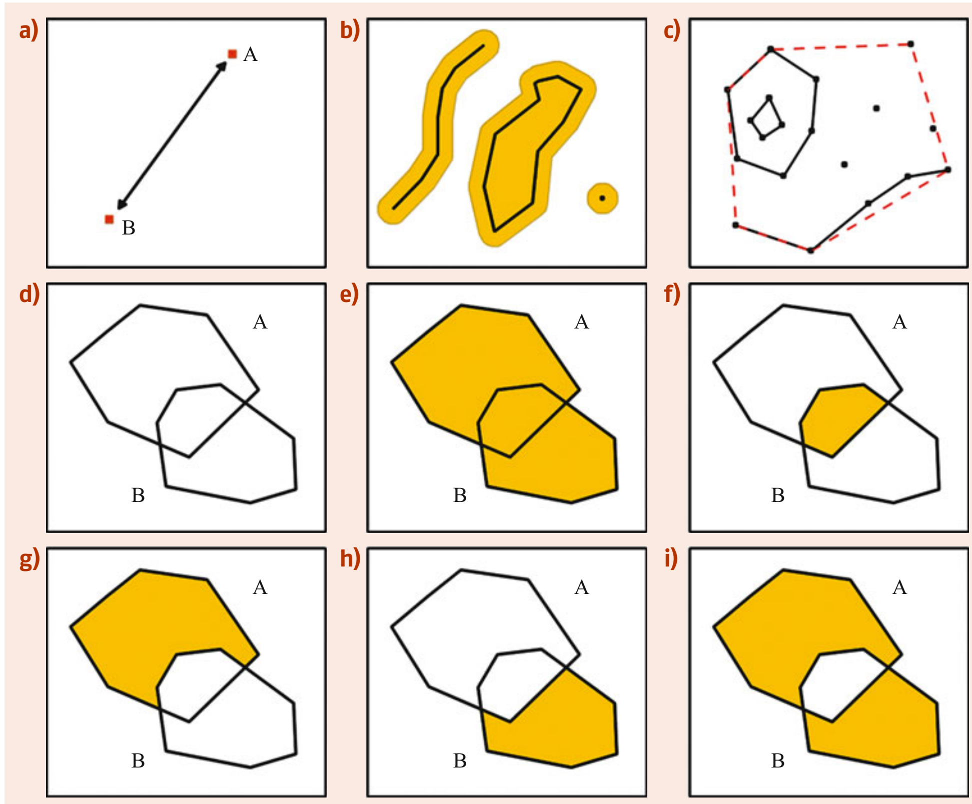


Fig. 30.8 Spatial functions supported by PostGIS: Distance (**a**), Buffer (**b**), Convex Hull (**c**), Union (**e**), Intersection (**f**), Difference (**g**,**h**), Symmetric Difference (**i**); **d** shows the polygons used for the spatial operations of (**e**)–(i); **g** shows the difference of polygon A to polygon B; **h** shows the difference of polygon B to polygon A

an open source and OGC compliant solution. Thus, PostGIS is supported by many Open Source GIS applications (OGR/GDAL, GRASS GIS, QGIS, MapServer, GeoServer, GeoTools, gvSIG) and also proprietary software like ArcGIS from Esri or FME from Safe Software.

```
SELECT DISTINCT a.city_name
        FROM city a, river b
                WHERE ST_DISTANCE (a.geom, b.geom) < 50000
                AND b.river_name = 'Isar';

                city _ name
------------------------------------------------------
Munich
Passau
…
```

Fig. 30.9 A spatial SQL query for the names of cities located within a distance of 50 km to the river Isar in Bavaria, Germany, provides the results for Munich and Passau

30.4.2 rasdaman

While pixels have traditionally been aligned horizontally in 2-D, this is just a special case; generally, observed or simulated pixel and voxel sets can be aligned in space and time, resembling 1-D time series, 2-D images, 3-D *x*/*y*/*t* image time series and *x*/*y*/*z* geophysical data, 4-D *x*/*y*/*z*/*t* atmospheric and ocean data, etc. With rasdaman ("raster data manager") [84, 89], such multidimensional raster data can be managed, accessed, and analyzed flexibly in a scalable manner. Based on a high-level raster query language users can ask any query, any time, on any volume. Hence, rasdaman resembles an Array Database System [90] or, in recent terminology, a datacube engine [91].

Historically, with rasdaman, datacube services were invented [92, 93]. Its early patents have retained openness and prevented later patents in the field to reserve the field for large players like Oracle. In fact, rasdaman has pioneered the field of Array Databases and datacube services in particular in science and engineering, with a large, growing number of epigons appearing recently. In 2008/2009, the keeper of the rasdaman code, research spinoff rasdaman GmbH, forked open source *rasdaman community,* which now is maintained by the company together with Jacobs University. The client libraries are licensed under LGPL so as to not "infect" any tool utilizing rasdaman while the server (running as a set of standalone demon processes) follows a GPL regime, based on the idea that life is a give and take. Today, the rasdaman community is used by science sites and companies like EOfarm/Greece, sees about 15–20 downloads daily, measured on average [89], and is official Open Geospatial Consortium (OGC) WCS (i.e., datacube API) reference implementation.

The rasdaman architecture consists of a full-stack implementation from scratch, with every component optimized towards large *n*-D arrays. Adaptive partitioning al-

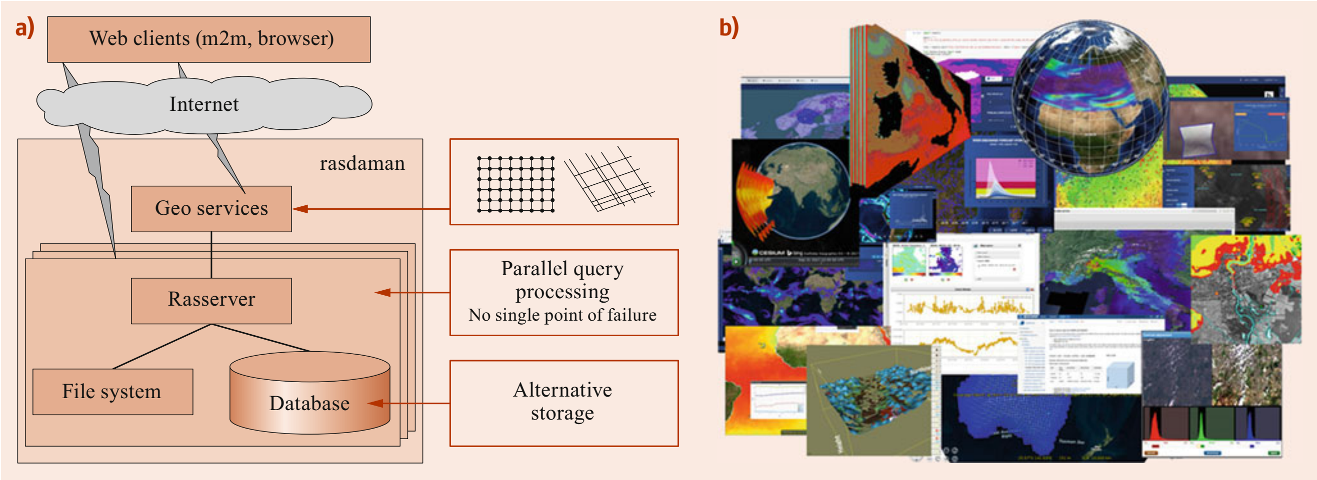


Fig. 30.10 rasdaman architecture (**a**) and kaleidoscope of rasdaman-enabled datacube portals (**b**)

lows tuning the database for optimal access performance (Fig. 30.10) [94]. Processing is multiparallel and scales vertically as well as horizontally. The engine as such is domain-neutral and has been utilized, for example, on human brain imagery, gene expression research, and cosmological simulation. An application layer [95] adds geosemantics, offered through OGC WMS, WCS, WCPS, and WPS interfaces, thereby understanding space and time and regular as well as irregular grids.

Thanks to these standards, rasdaman can interface with a large number of clients, which allows users to stay in the comfort zone of their well-known tools, including OpenLayers, Leaflet, QGIS, ArcGIS, NASA WebWorldWind, Cesium, C++, Java, python, and R. However, interfaces to other server tools exist as well, such as MapServer and THREDDS, extending these with flexible, scalable raster capabilities.

The concepts of rasdaman have massively impacted standardization. ISO Array SQL [26, 96, 97] adds rasdaman's datacube modeling and querying to the SQL language in a domain-independent manner. The OGC datacube standards suite, WCS, and in particular the OGC datacube analytics language, WCPS, was crafted in different syntax but with the same operators, extended with spatiotemporal semantics [98] (see [99] for an introduction to these OGC standards). While crafting the multidimensional datacube standards, the rasdaman team also had to invent APIs for spatiotemporal coordinate reference systems [100] and today operates OGC's Coordinate Reference Systems resolver.

30.4.3 GeoBlacklight

Information search and retrieval engines are increasingly reliant upon spatial data infrastructures (SDIs) in order to manage the querying, access, and visualization of geographic information. While the components and scale of one single SDI implementation will largely depend on the specific needs of its primary users, geoportals have emerged as a key component of an SDI as a means to search and retrieve content. Historically, discovery and reuse of geospatial data was constrained by numerous issues involving metadata, infrastructure longevity, persistent data access, and a general lack of simplicity in search interfaces. To address some of these challenges, partner institutions at MIT Libraries, Princeton University, and Stanford University Libraries developed GeoBlacklight, a discovery platform for geospatial data.

GeoBlacklight is an open source collaborative software project designed to facilitate the finding and sharing of geographic data [101]. This project builds upon the previous work of the Blacklight and OpenGeoportal communities, primarily in terms of its approaches to interface design and metadata strategy. Built on top of Blacklight and Ruby on Rails, GeoBlacklight provides a discovery engine for geospatial data. It integrates and uses existing web standards, including Open Geospatial Consortium (OGC) Web Map Service (WMS), Web Feature Service (WFS), GeoJSON, and the International Image Interoperability Framework (IIIF). Since its initial release, several partners have adopted the software and are actively contributing to its ongoing development and enhancements [2]

GeoBlacklight and its underlying metadata schema support common use cases in data discovery for search, view, and curation activities [102]. Discovery support is enacted at the layer level, meaning that each specific unit of data containing spatial features (shapefile, GeoTIFF, digital map, etc.) can be visualized, downloaded, and cited. Search options include spatial map-based search, natural language text-based search, and faceted refinement of result sets. These three search modalities mimic modern everyday user interactions with the web using a single search bar, dynamically updating maps, and categorical results filtering.

GeoBlacklight is willfully agnostic towards institutional metadata operations. This is an intentional design decision due to the vast array of standards used to describe spatial information, ranging from ISO (for GIS data) to MARC and MODS (for paper and digital maps). Prioritizing one metadata standard over another could limit widespread adoption of the software as well as the data sharing capabilities. Furthermore, these standards are unnecessarily complex for the purposes of discovery.

The GeoBlacklight Metadata Schema is a set of elements standardized for the discovery of geospatial data. Formatted in JSON, the schema is an extension to the Dublin Core and DCMI Terms vocabulary. It uses GeoRSS semantics to encode geospatial features such as bounding boxes, points, and polygons (expressed as WGS84). This relatively simple schema is designed to enable the normalization and interoperability of metadata contributed from a variety of data providers. Organizations can create and manage records natively in the GeoBlacklight schema or automate transformation of their metadata from other standards.

The schema requires a small number of fields, which are crucial for basic functionality (e.g., bounding box, title, and identifier). Most elements are optional, including citation details (e.g., creator, publisher, and publication date) and subjects of the data (e.g., geographic names, topics, and temporal extents). The element dct_references uses JSON-LD syntax to link to data downloads, external services (WMS/WFS/WCS), codebook documentation, and additional metadata formats. Links can also incorporate schema.org tags to make data layers discoverable through web search engines.

OpenGeoMetadata is a multi-institutional GitHub repository of spatial metadata [103]. Contributing organizations manage their own repositories in which they store GeoBlacklight JSON as well as metadata in any other format. This allows for the harvesting of multi-institutional JSON metadata within a localized GeoBlacklight instance. The aggregation of metadata will often expose divergences in local cataloging practices with regards to quality and the use of controlled terms to represent places, people, or things [104]. These issues are not specific to geospatial data. The OpenGeoPortal and GeoBlacklight user communities are continually investigating methods to develop best practices and encoding standards for multi-institutional geospatial metadata in order to address quality, record completeness, and the interoperability of metadata in a shared data catalogue.

30.5 Spatial Tools

Specific analysis tools exist for a wide range of application fields. In this section, we focus on cartography and digital field mapping.

30.5.1 The Generic Mapping Tools (GMT)

The Generic Mapping Tools (GMT; https://www.generic-mapping-tools.org) is widely used for analyzing and displaying a wide variety of data [105–109]. In particular, its capacity for analyzing and processing geoscience data and creating publication-quality graphics has made it an important toolset used by the Earth, ocean, and even planetary sciences communities. GMT's strengths lie in its flexible vector graphics (from page-size to wall-size), geodetic-quality map projections, and robust data processing algorithms scalable to enormous datasets and grid sizes, e.g., make global bathymetry maps from 440 million randomly-distributed soundings constraining grids with sizes up to 43 200 by 86 400 [110]. GMT runs under all operating systems, is Open Source and freely available under a flexible license (Lesser GNU Public License), and it offers $\approx$140 modules sharing a common set of options, file structures, and documentation. GMT modules are implemented as Unix filters, which let users write scripts where one module's output may become another module's input, creating customized GMT workflows. GMT depends on netCDF and GDAL for handling a variety of input and output file formats, and use libcurl and libpcre for remote file access and pattern matching. It installs with the global coastline, rivers, and political border database GSHHG [111], allowing for simple basemaps to be generated (Fig. 30.11).

Since its inceptions over 30 years ago, GMT has remained a command-line tool set and has greatly benefitted from scripting [10]. However, the release of GMT 5 [112] introduced three key changes: (1) a documented C Application Program Interface (API) was added, containing basic functionality for handling GMT data objects (i.e., the input/output of data), manipulating module options, reporting errors and warnings, and accessing any of $\approx$140 modules. In GMT 5, these modules are no longer standalone programs but have been reimplemented as high-level API functions; (2) to avoid "namespace pollution", a single executable called *gmt* was build, that can access all modules; (3) the concepts of input sources and output destinations have been generalized. In addition to the passing of file names or using standard input/output, developers using the API can select sources and destinations in several different ways, including memory objects, file pointers to open files or standard streams, and file descriptors. The GMT modules themselves

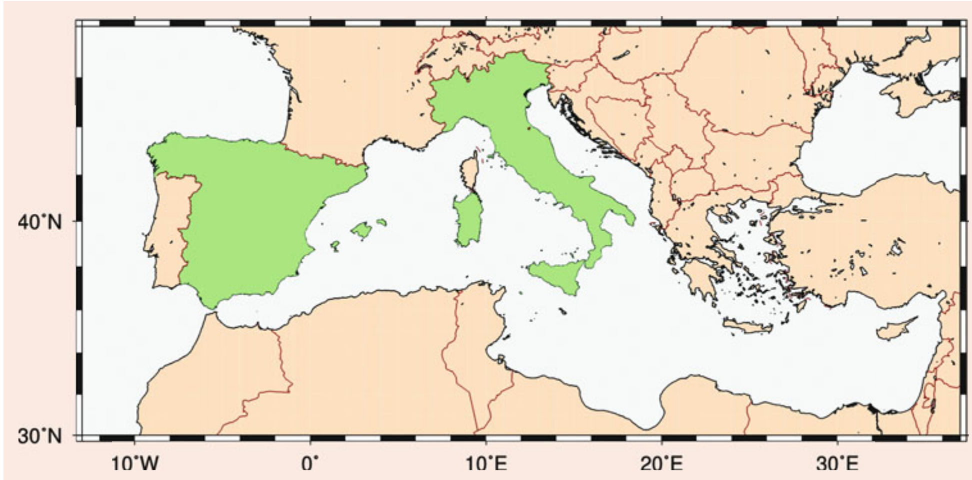


Fig. 30.11 Using built-in data we can make a simple basemap of the Mediterranean countries, highlighting Spain and Italy

are largely unaware of this, as the flexibility is implemented in the API input/output layer. Hence, the GMT API makes it possible for developers to build shared libraries to access custom modules on top of the core C API. One such external shared library extends GMT to handle specific processing for oceanic fracture zone and magnetic lineation data [113].

Going forwards, new and existing GMT users can choose to access GMT modules from their favorite scripting environments. The shell/DOS scripting use is the oldest and most established GMT. This command line syntax is called the *classic* mode, but in GMT 6 we introduced a new, improved and simpler scripting interface (called *modern* mode and published on the open-access repository Zenodo) [105, 114]. In addition, new interfaces to GMT are being designed rapidly (Fig. 30.12). A GMT toolbox for MATLAB and Octave has been released [115]. and both Python [116] and Julia [117] interfaces are steadily improving. The latter two simplify the interface by using keyword-value arguments instead of the traditional letter options used on the GMT command line. Similar to how GMT shell scripts take advantage of other Unix tools, scripts written in MATLAB, Python, and Julia can take advantage of their computational frameworks to achieve shorter, faster, and more customizable results. This "best of both worlds" approach will lead to a shorter learning curve for GMT: Both Python and Julia can access the GMT API from environments such as Jupyter notebooks and, thus, be more interactive. Coupled with the access to remote datasets, it is now easier than ever to make more sophisticated maps or animations, such as requiring global relief files (Fig. 30.13).

All repositories related to the Generic Mapping Tools are available on GitHub, with the address https://github.com/GenericMappingTools.

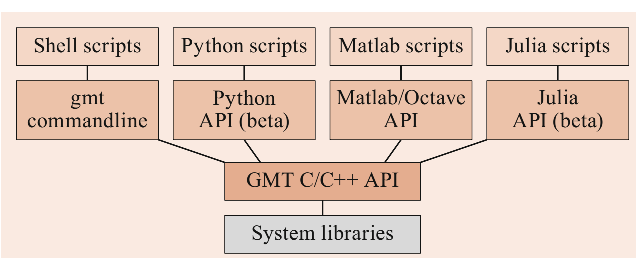


Fig. 30.12 Block diagram of GMT accessibility. High-level functionality resides in the API, and modules are called via the command-line *gmt* executable. Custom APIs may also be accessed this way. The GMT/-MATLAB, GMT/Python, and GMT/Julia APIs offer scripts written in those languages direct access to all GMT modules

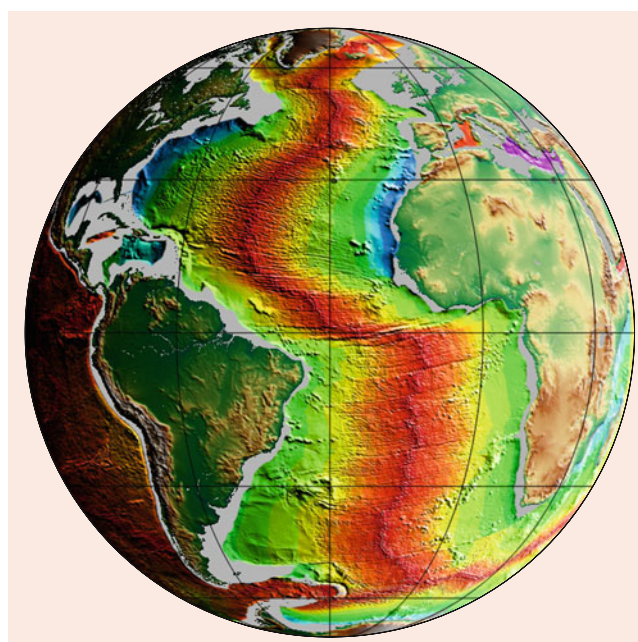


Fig. 30.13 Land relief [118] and ocean crustal ages [119], shaded by the global relief, with ambient light modulated by GMT to simulate sunlight

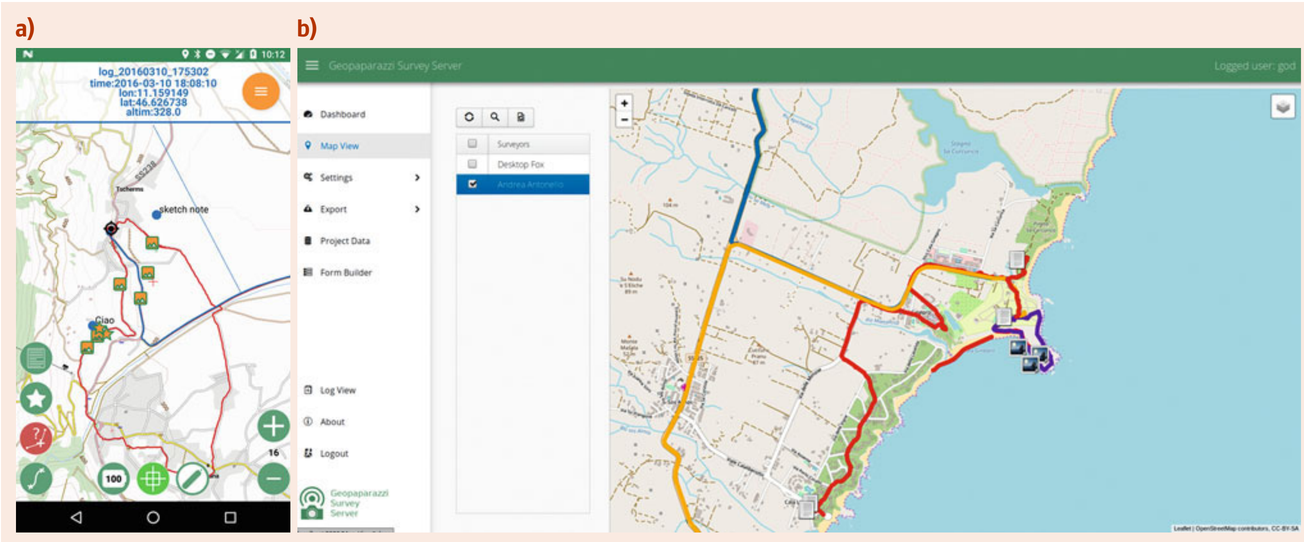


Fig. 30.14 **a** The map view of Geopaparazzi with the log analysis tool activated. In the view, the GPS position icon is visible, as well as various recorded GPS logs, text and images notes, and bookmarks. Location: Tscherms, Italy. **b** Map view of the Geopaparazzi Survey Server, visualizing the datasets synchronized by one of the two registered surveyors. Location: Cala Liberotto, Italy

30.5.2 Digital Field Mapping: Geopaparazzi

The digital field mapping application Geopaparazzi [120] is not the first open source application developed for digital field mapping. Already around 2004, developers installed the GIS GRASS on early PDAs like iPAQ and ZAURUS [121]. To get an idea of the computational power of these devices, the iPAQ (model 3870) used was equipped with an Intel StrongARM SA-1110 206 MHz processor and had 64 MB of RAM. The result was a proof of concept, not really usable in the field. However, the desire for digital field mapping applications on lightweight devices was huge. In some environments, customizations of complete GIS applications were used on rugged tablet with a complete windows OS to provide pen-based data management, as well as proper GPS integration. One example of such an application is the open source GIS for field mapping BeeGIS [122]. However, the mobile trend moved quickly from rugged tablet pcs and ultramobile pcs to smartphones and tablets. Loading a complete GIS on a low-power and small screen device that needed to be handled in the field with a pen or fingers did not result in an efficient tool whose objective was to substitute the use of paper (an example is the geological field book) allowing direct digitalization, hence removing all the errors that appear during the data postprocessing previously done in the back-office.

The first successful open source application on small portable devices was gvSIG Mobile, an application created for Windows CE and often used combined with Trimble GPS devices. The application gained a lot of traction and was in use for several years, even after its development had stopped, because a real open source alternative was missing. The real problem of gvSIG Mobile was the timing in the choice of the operating system (OS). In 2010, when gvSIG Mobile version 0.3 was released, Windows CE was a dying OS, and in that timeframe the two new major players of the mobile world were defining their positions: IOS and Android. By the end of 2010, the mobile market share was distributed between Symbian and RIM, former market leaders (both already in steady loss) and the fast growing Android and IOS operating systems [123]. This created a new mobile market that produced different lines of products, making smartphones available at reasonable prices. Also, the outdoor screen visibility and the battery lasting time experienced a huge growth, making these smart devices the perfect tools for digital field mapping.

In 2010, the first version of Geopaparazzi, an app for digital field mapping on Android devices, was released. At the time, the app had few but well-defined features. It allowed recording simple text notes, capturing pictures with the information of the snap direction, and recording GPS tracks. It never aimed to be a mobile GIS, instead it put its focus on the ease of use and robustness of its features. At the time of writing this book chapter, Geopaparazzi had reached version 6.0.4. Its main features are:

- A main dashboard view to access the main functionalities
- A map view with navigation tools and tools for qualitative distance measurement, analysis of the GPS logs, and handling of notes and bookmarks
- Recording of GPS tracks, with the additional possibility to chart the tracks visualizing elevation and speed

- Capture of different types of notes: simple text, image notes, sketch notes, and complex form based notes (Fig. 30.14a)
- Support for visualization and editing of spatialite vector databases. Spatialite is a library that extends the SQLite database to support spatial SQL capabilities. Spatialite is present on all major operating systems, such as Windows, Linux, Android, and IOS [124]
- Import and export functionalities, mainly for GPX and KMZ formats, as well as through connections with cloud servers
- Support for background maps, being file-based or online (OGC TMS, WMS standards). The mapsforge project provides a free map rendering library. It was started in 2008 at the Free University of Berlin and targets the Android platform [125]. MBTiles is a format to store tilesets. It is based on an SQLite database and its specifications are open [126].

In 2017, the gvSIG Association decided to choose Geopaparazzi as the base for the mobile part of its software stack, gvSIG Mobile [3]. This event triggered interest from various administrations and enterprises, which resulted in the need for data centralization in the cloud. Different companies started the development of server applications that would be able to interface with Geopaparazzi and gvSIG Mobile. The most important are:

- The Geopaparazzi Survey Server (GSS) [127]: a web application in which surveyors can be registered through the unique device id. The app Geopaparazzi Survey Server Sync is provided to allows user to synchronize a Geopaparazzi project, download project data from the server, and download form definitions that have been created on the web. On the server, users can navigate and query the synchronized datasets in a map view, export datasets to PDF or KMZ, upload project data and interactively build complex forms that are then made available to the client Android application (Fig. 30.14b).
- gvSIG Online [128]: a web application that allows to connect to gvSIG Mobile and synchronize spatialite vector datasets. In the web application, the user can select a set of vector layers. These layers are then made available to gvSIG Mobile, where they can be edited and finally reloaded to the server.
- The Cloud Profiles Server [129]: a web application that lets Geopaparazzi download profiles directly from the import menu. Profiles allow the user to download a complete survey configuration from the server. Survey configurations can provide a project and datasets and form definitions, putting the surveyor in a ready-to-go status without any need for prior knowledge about the location of all the data needed for the survey.

Fig. 30.15 The main view of S.M.A.S.H. visualizing a PostGIS polygon layer styled using OGC SLD theming

In 2019, an effort to make Geopaparazzi available on both Android and IOS devices began under the name S.M.A.S.H. (the Smart Mobile App for Surveyor's Happiness) [130]. After two years that effort had produced a rewritten, more modern, userfriendly and featurerich Geopaparazzi. All features of Geopaparazzi, apart of 3D and SpatiaLite support, had been replicated. And several new features had been added, the most notables being: PostGIS and GeoPackage read/write support, shapefile read support, GeoTIFF support, OGC SLD vector styling, Kalman filtering for GPS logs [4].

30.5.3 GeoPython

GeoPython is not a fixed term but is well known in the scene dealing with Free Software and GIS (FOSSGIS, FOSS4G). In the most general way, it is used for solving geospatial problems with Python at all. Since 2016, even a conference named GeoPython has been held every year in Basel [131]. In recent years, this GeoPython way has been very popular. Therefore, many good reasons are related mainly to Python itself [132].

1. Python is available for most common operating systems, mainly LINUX, Mac OS, and Windows, and it is foremost use cases make it easily possible to transfer the code base from one system to another.
2. Python is a very logical and structured programming language.
3. Python comes with many libraries on board. This circumstance is often also referenced as "batteries included" and describes that the standard library covers nearly everything from audio support and asynchronous processing to zip files.
4. Python is the most popular language for data science. It provides many libraries, like NumPy, SciPy, Pandas,

or Matplotlib for advanced mathematical analysis, which supports nearly every task in the broad field of data science.
5. Python is also the most popular language for questions that can be answered by artificial intelligence. This includes supervised learning, unsupervised learning, neural networks, deep learning, and much more.

In recent years, Python has also become the most prevalent language for covering geospatial tasks. There are libraries for nearly every geo-related problem, and the term GeoPython is used in a narrower sense for all Python libraries that are related to geoinformatics, e.g., rasterio, fiona, shapely, etc., and also for geo-related software written in Python, e.g., PyCSW [5], PyWPS [6], GeoHealthCheck [7], etc. In addition to these, APIs are available for the most popular GIS projects. These include APIs for free software like GDAL/OGR [133], PROJ.4 [134], GRASS GIS [36], QGIS [135], but also APIs for their proprietary counterparts like Esri ArcGIS or FME from SAFE Software. The following descriptions are from Internet sources for some important GeoPython libraries and give only a very brief overview of their functionality.

- *shapely* is a Python library for manipulating and analyzing planar geometric objects. It is based on the well-known GEOS and JTS libraries. GEOS is the C-based engine of PostGIS, and JTS is the Java library from which GEOS is ported. shapely is not concerned with data formats or coordinate systems but can be readily integrated with packages handling this task. shapely is BSD licensed [136].
- *rasterio* provides access to geospatial raster data (satellite data, digital elevation data). rasterio reads and writes raster data formats and offers a Python API, which is based on Numpy N-dimensional arrays and GeoJSON [137].
- *fiona* focuses on reading and writing data in standard Python IO style and wraps the OGR library. Therefore, it provides familiar Python types and protocols such as files, dictionaries, mappings, and iterators instead of the specific OGR classes. fiona can read and write spatial data from multilayered GIS formats to zipped virtual file systems [138].
- *PyProj* is a Cython wrapper to provide python interfaces to PROJ.4 functions. It performs cartographic transformations and geodetic computations [139].
- *GeoPandas* is an open source project to enable working with geospatial data. According to this, GeoPandas extends the data types used by pandas to allow spatial operations on geometric types. These geometric operations are performed by shapely [140].
- *folium* transfers the Leaflet.js functions from JavaScript to Python and enables the Python developer to create maps. Folium brings the data power of the Python ecosystem and the mapping strengths of the Leaflet.js library together in one application. So, it is possible to manipulate your data in Python and visualize it in a Leaflet map via folium [141].

The really big advantage of following this GeoPython approach is that the geoscientist is able to deal with the spatial data in the way he/she is used to from his/her known GIS software. On the other hand, he/she has all the great variety of capabilities offered by Python. Among others, these range from automating tasks over scientific programming to machine learning. For example, it is possible to load satellite data with GDAL or rasterio and stack the band data directly to a NumPy ndarray or an xarray. This xarray can be converted to a Pandas object or loaded to machine learning libraries to perform further analysis or classifying the image. The result can be plotted with folium and Jupyter Notebook or Basemap and Matplotlib (Fig. 30.16). Especially in the area of remote sensing huge amounts of datasets have been available over recent years, so that it is no longer convenient to download the satellite data and process them locally. Instead, the data are stored within large processing platforms, and the processing is performed remotely. For this approach, the Jupyter Notebook is an extremely convenient solution. The Jupyter Notebook App is a server–client application. This application enables users to edit and run Notebook documents via a web browser. It can be installed on a remote server and accessed through the Internet [142]. Jupyter Notebook provides the same functions via HTTP as any other Python shell. Examples of such usages are notebooks available from CODE-DE [143] and also from Planet Labs [144]. Besides this, the Jupyter Notebook App can be executed on a local desktop without any Internet access. In this case, it is extremely useful for teaching and holding workshops. With JupyterCon, also a conference around the Jupyter universe exists.

30.6 Web Services

Web services are used to publish spatial data via the World Wide Web, adhering to established standards such as AJAX or WxS by OGC. They typically provide web-based user interfaces, which are provided through web browsers. Web services can both offer and consume geospatial content.

30.6.1 MapServer

The University of Minnesota MapServer project (henceforth referred to as *MapServer*) emerged from National Aeronautics and Space Administration (NASA) sponsored research

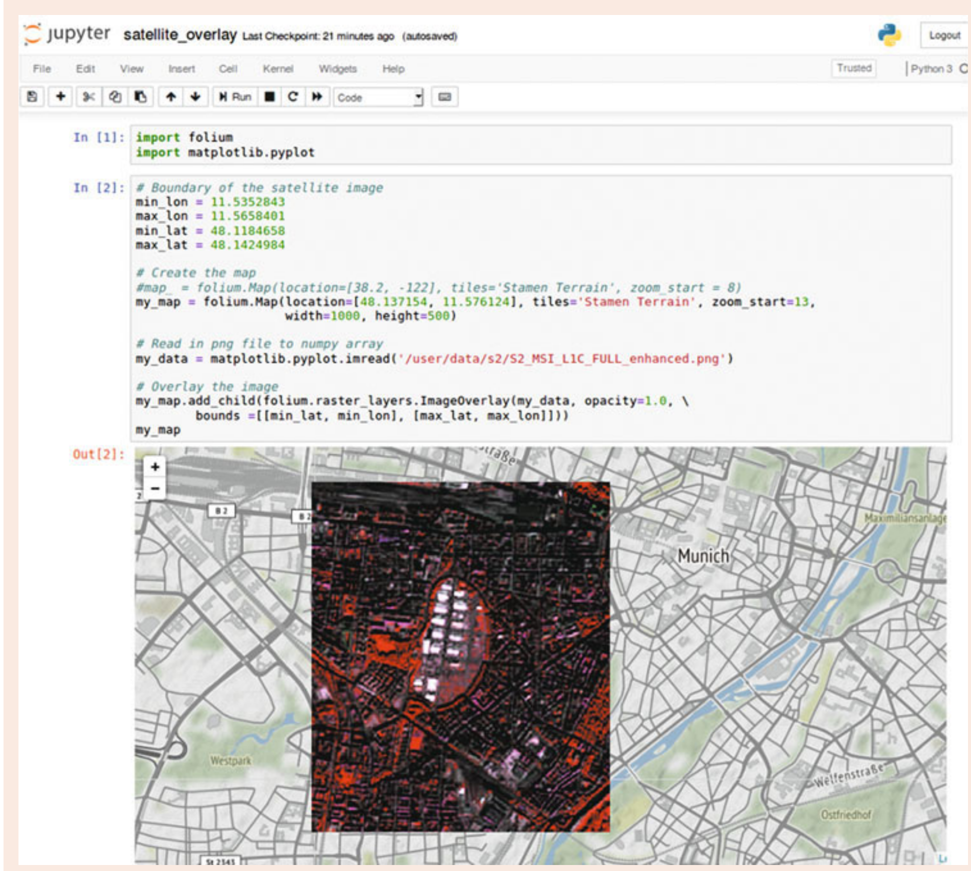


Fig. 30.16 Cartography with Jupyter Notebook and folium. This example shows a satellite overlay of Sentinel-2 imagery for the Theresienwiese (Oktoberfest venue) in front of an interactive topographic map of Munich, Germany

at the University of Minnesota, starting in 1994, to become one of the most popular Open Source GIS packages available today. MapServer integrates with other popular Open Source applications such as PostGIS, Leaflet and OpenLayers and supports most major Open Geospatial Consortium (OGC) Web service specifications. MapServer is a platform for developing spatially-enabled, web-based applications and services [19]. The software, written in ANSI C, runs on just about any computing operating system, from AIX to Microsoft Windows. MapServer (as of this writing the current version is 7.6.4 [145]) comes with an out-of-the-box common gateway interface (CGI) application that provides functionality to build interactive websites and Open Geospatial Consortium (OGC) service instances.

The broader MapServer project is host to other related projects – most notably MapCache [20] – a high-performance tiling engine.

Project History and Relation to OSGeo and Repository Infrastructure

MapServer is different from many open source geospatial projects because it was not the product of a formalized development process; there was no conscious effort to build what exists today. On the contrary, MapServer is the product of countless hours of prototyping, of writing and rewriting, and trial and error, especially early on. Instead, functionality evolved organically. When something new was needed it was simply added, either via new code or by integrating an external library.

MapServer was first licensed under an open Source MIT/X11 license in 1999 (MapServer 3.0 release).

MapServer transitioned to OSGeo infrastructure, e.g., SVN, in 2006 and formally graduated from the OSGeo incubation process in 2008. OSGeo became the formal copyright holder in 2008.

MapServer eventually moved off OSGeo application development infrastructure to GitHub [146]. As a platform GitHub allowed a new generation of developers to contribute to the project with amazing ease by code forking. MapServer has always stood on the shoulders of best-of-breed open source libraries and continues to do so on the infrastructure side as well.

OGC WxS Specification Support

The support of Open Geospatial Consortium (OGC) Web service specifications (WxS) by MapServer has grown continually. The major WxS service specifications map directly to major MapServer functional areas.

WMS support (for both client and server) was introduced in MapServer version 3.5. Recently, WMS versions 1.1.1,

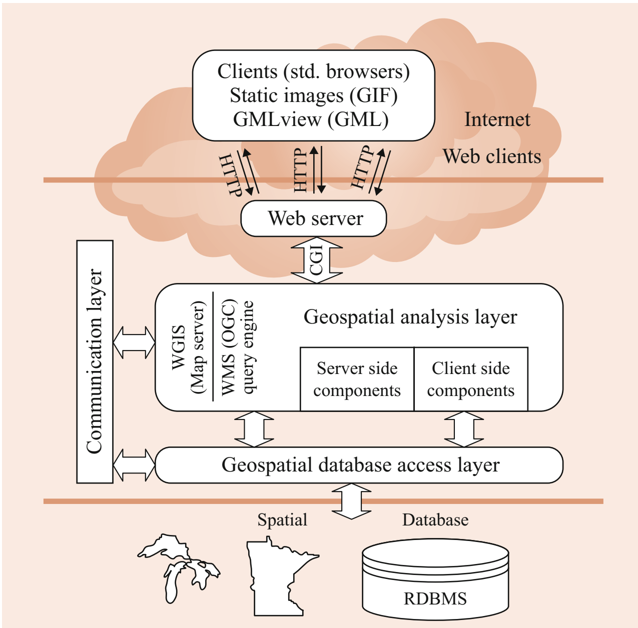


Fig. 30.17 High level view of the MapServer software architecture

1.2, and, most recently, 1.3 are supported. The useful, but little used, Web Mapping Context (WMC) was supported beginning with MapServer 3.7.

WFS 1.0.0 support (also for both client and server) was included with the release of MapServer 4.0 with support for WFS 1.1.0 and 2.0.0 added subsequently. WCS 1.0.0 support (server only) has been supported since MapServer version 4.4.

WFS-T, WMS-T, and TMS have also been supported through the MapCache and TinyOWs libraries since 2012. The most recent addition to MapServer is for the Sensor Observation Service (SOS) specification, which was added at version 4.10. In all cases, the MapServer CGI interface is used to front the OGC service implementations (Fig. 30.17).

Map Elements and Output Formats

MapServer supports the automatic generation of components that a user would normally find on a map, including legends, scale bars, and reference or key maps. In addition, these elements (primarily scale bars and reference maps) can exist independently from one another so that, for example, MapServer can be used only as a scale bar engine. While at its core MapServer makes maps, it also contains a reasonably powerful feature query capability. Originally, MapServer only supported point selection functions. Over time, additional variations have been added, including query by attribute and by arbitrary polygon as primary examples of these.

While HTML is used most frequently as a template, it is possible to create other types of output such as plain XML, KML, SVG, PDF, UTMgrid, GeoRSS, or JSON.

Mapfiles

Configuration of a MapServer application consists of a single text file called a *mapfile*. A mapfile has a hierarchical format consisting of a series of object definitions that describe an application or service, the data to be used, presentation definitions for the data, and, finally, additional elements that may be part of a service or application, e.g., scalebars, legends, and reference maps. The structure of a mapfile is comparable to XML. It consists of tags or identifiers to start object definitions and the keyword END to close an object. Comments are denoted with a # and quotation marks are used to delineate data containing special characters.

The example of a mapfile for a polygon layer is provided in in the mapfile MST1. Adding additional data to the map is a matter of adding additional layer definitions to the mapfile. The order of certain types of objects is significant. For example, layers are drawn in the order that they appear in the file. So raster layers are typically added first, polygon data next, and, finally, line work, points, and annotation.

Mapfile MST1: Sample mapfile defining a 500×500-pixel map called myMap, containing a single polygon layer drawn with a solid red fill from an Esri Shapefile as data source

```
# Simple Mapfile with a complete definition
MAP
  NAME 'myMap'
  EXTENT 50000 100000 60000 101000
  SIZE 500 500
  UNITS METERS
  LAYER
    NAME 'myLayer'
     DATA 'myLayerData' # no .shp extension
    TYPE POLYGON
    STATUS ON
    CLASS
      STYLE
        COLOR 255 0 0
      END
    END
  END
END
```

MapServer mapfiles are decidedly home-grown and have evolved ahead of technologies such as XML or JSON. XML support was eventually added in version 6.x where XSLT is used to transform an XML configuration into a standard mapfile format. Other interesting configuration alternatives exist as well, including the MapBox GL style specification, in JSON.

Configuration format aside, a WYSIWYG (what you see is what you get) editor for mapfiles has proven to be an elusive target.

Any number of external projects, including MapManager, ScribeUI, and MagnaCarto, are still available (in various stages of development). Extensions to ArcView, gvSIG, and QGIS attempted to leverage popular desktop GIS tools to author mapfiles. However, differences in symbology models often make this a difficult task. QGIS paired with the useful mappyfile Python library represent a particular interesting approach, as it allows to import a MapServer mapfile in a QGIS project.

The concept of a *layer* is central to most GIS, and MapServer is no different. A layer holds all the information necessary for the software to access, query, and render features. In most cases, a layer refers to a GIS dataset that will be drawn in an image buffer. Data can come from external file-based sources, e.g., Shapefiles or TIFF images, a spatially enabled database, e.g., PostGIS, OGC services, e.g., WMS or WFS, or even features defined within the mapfile itself.

To determine which presentation definition, e.g., color, symbol, size, is required, a "class" object is used. All layers being drawn must have at least one class defined. Classes store map presentation definitions in style and label objects and query output definitions in template references. Most importantly, classes are home to expressions that are used to group features logically for drawing.

Style and label objects contain the actual definitions used by MapServer rendering functions to color pixels in a map. Styles control which symbol and what color to use, and how big to make it, while labels control how any annotation associated with a feature will look, including font, size, color, etc.

A class can contain several styles, so in effect, styles can be stacked to produce more complex cartographic output.

MapServer supports the concept of runtime geometry transformations so that different representations, ntroid, buffer, and first/last point, can be used for styling or labeling. This allows for complex cartography, such as drawing arrows at the end of a line or highlighting the bounding box of a specific polygon.

The map object also holds definitions for other objects that might normally be seen as part of a map or a mapping application. These consist of scale bars, legends, and reference maps.

MapServer legends take much of their configuration from the layers, classes, and styles described above. It consists primarily of parameters to control the size and spacing of key boxes and any fonts that might be used. MapServer supports two types of legends, namely an image-only version that, like scale bars, can be standalone or embedded and a text legend. Text legends most often take the form of HTML and are produced using a legend template.

Reference maps are small "locator" maps that are intended to help provide a frame of reference when a user is zoomed into an area. These are created from a predefined base graphic with a known size and extent (often produced by MapServer). Rectangles or markers representing the location of another map are then rendered on top of the base map.

MapServer supports five basic types of symbols:

1. Pixmap symbols: images used as markers, brushes or fill patterns.
2. Vector symbols: inline coordinate sequences used to render markers, brushes, or fill patterns. They have the advantage of being scalable.
3. Ellipse symbols: ellipses (most often just a circle) rendered into a marker or a brush. They are most often used to draw wide lines.
4. TrueType symbols: similar to vector symbols except that the coordinate information is held in a font glyph.
5. Hatch symbols: a special polygon fill that supports arbitrary hatch angles. There are a number of secondary parameters such as anti-aliasing, fills, and dash patterns that can be applied as necessary.

Mapscript and Scripting Languages

MapServer also features MapScript, a powerful scripting interface for popular languages such as PHP, C#, Java, Perl, Python, and Ruby. MapScript allows developers to add geospatial functions to any application.

The easiest way to understand MapScript is to realize that with it a developer could rewrite the out-of-the-box CGI interface with it. The CGI is written on top of the C API of MapServer, as is MapScript. For example, reconsider Fig. 30.17. MapScript can access functions in any of the boxes (with the exception of the CGI interface). All of the relevant drawing, query, data access, and presentation methods are available to the developer.

Scripting languages like Perl, PHP, Python, Ruby, Java, and C# allow to expose MapServer core functionality to a large community of application developers. The true power of scripting languages is that developers have many other toolsets at their fingertips. There are modules for everything, so it becomes trivial, for example, to retrieve a web page, parse a coordinate from the content, buffer the coordinates, query a data layer with the resulting polygons, and map the results.

Current Role and Future Challenges

It is important to realize that MapServer is part of the larger FOSS4G stack, of which each component has strengths and weaknesses. MapServer's role is clearly on the server side in the production of maps and data in support any number of client environments. Over time, client-side application development systems including Leaflet, OpenLayers, and GeoMoose have emerged, which use MapServer and/or OGC services. They have formed the basis for the development of any number of web applications.

As new standards or de-facto standards emerge, e.g., KML, GeoJSON or GeoRSS, and MapServer should be an early adopter whenever possible. Ultimately improving

MapServer support for style layer descriptor (SLD) will go a long way towards cartographic interoperability. This should be a focus of future MapServer development.

30.6.2 GeoServer

GeoServer is an open source, server-based software that enables publishing and editing of geospatial data through geospatial standards [8, 147, 148]. The software implements Open Geospatial Consortium (OGC) standards, including Web Map Service (WMS) [149], Web Feature Service (WFS) [150], and Web Coverage Service (WCS) [151], and is used broadly. Used as a platform for a national spatial data infrastructure (SDI) or custom research projects, GeoServer has a large and diverse body of adoptions worldwide.

GeoServer is a highly flexible software platform enabling broader access to many different types of data. GeoServer supports several connectors for both raster and vector data formats. This allows GeoServer to be plugged into existing technical architectures and/or ecosystems that prefer one approach over another. The software then publishes this data using OGC-compliant web services. Common use cases for GeoServer include serving imagery, historic maps, or vector data.

The standards implemented by GeoServer enable its data to be accessed and used by existing geospatial analysis and visualization software. Capabilities exist for integration into geospatial visualization software such as Google Maps. GeoServer is also flexible enough as a backend to be integrated with content management systems like GeoNode [152]. GeoNode uses GeoServer as a datastore and way to serve content.

Content in GeoServer can be managed in several ways. The GeoServer software provides a administrative interface where tasks, and settings can be managed. Also provided is a Representational State Transfer Application Programming Interface (REST API) [153] that can be used to manage administrative tasks and settings of GeoServer and upload data to be published.

Because GeoServer implements and provides access to data using geospatial standards, the software integrates and utilizes additional open source geospatial software. Included and used software include GDAL, OpenLayers, and GeoWebCache. GeoServer adopters can take advantage of use these projects in concert with GeoServer.

30.7 Geospatial Libraries

Software libraries implement frequently used functionalities, which are provided through well-defined interfaces (API) to be reused by multiple independent programs, which are not connected to each other. Libraries can themselves make use of other specialized software libraries in a manner of hierarchical dependencies. Libraries allow to cluster specific capabilities in reusable code and to focus the development effort and resources accordingly, instead of creating many redundant implementations among the independent programs that benefit from the libraries.

30.7.1 The Significance of Libraries for FOSS GIS: the MapServer Example

The advent of core geospatial libraries such as PROJ and GDAL/OGR has been essential for the development of FOSS GIS. This is exemplified here by the case of MapServer.

The MapServer project as it exists today was started by the release of Shapelib in the mid-1990s. When the MapServer drivers for spatial databases were written, a driver for the middleware library OGR was also added. OGR allows MapServer access to any data format that OGR can read, currently including over 29 formats.

Similarly, the first open source releases of MapServer had basic raster support by means of native drivers for a limited number of graphics formats. The raster translation library GDAL was added to MapServer at version 4.0 and allows access to around 60 different formats.

Clearly, it is not practical for a single project to develop all of these capabilities by itself. Fortunately, the OS world is replete with niche libraries, so that developing MapServer really became a discovery and integration task as much as anything else. MapServer relies on more than a dozen supporting low-level libraries and development tools. Without these there would be no MapServer. Hence, software libraries are the essence of open source development.

The MapServer project owes a debt of gratitude to the large collection of other open source projects that it leverages as libraries, including their dependencies. A graphical overview of the software libraries utilized by MapServer and their respective dependencies is shown in Fig. 30.18.

30.7.2 PROJ

PROJ is a generic geospatial coordinate transformation engine [155, 156]. PROJ supports "transformation pipelines" to concatenate simple coordinate operations into complex geodetic transformations of spatiotemporal coordinates. It is provided by the software project PROJ, which is currently scheduled for OSGeo incubation. The project name PROJ stems from the previous major version number (PROJ.4), which was released in 1993.

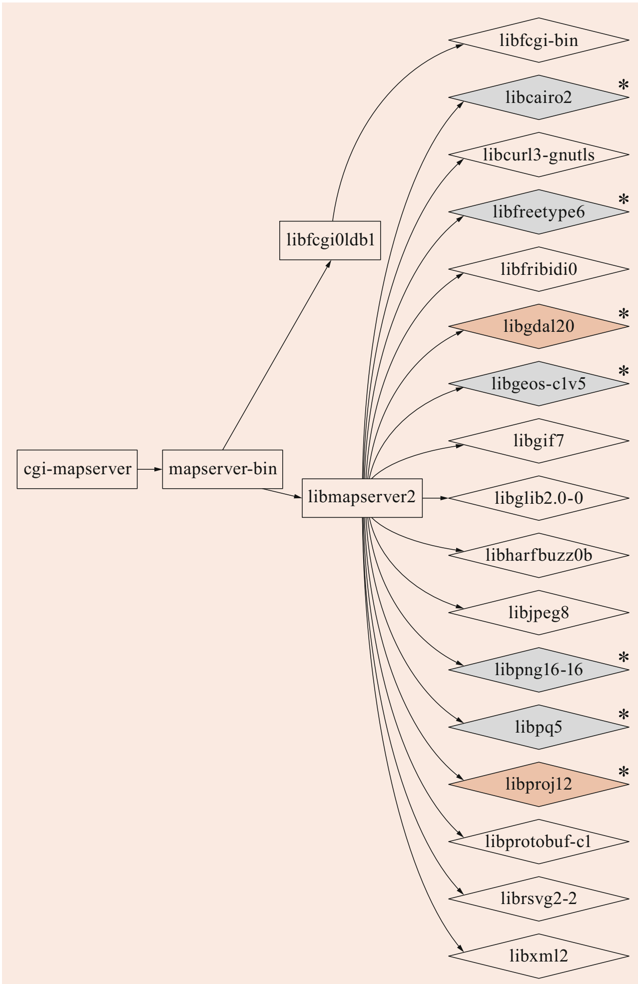


Fig. 30.18 MapServer software library dependencies overview [154] rendering: the first three layers of dependencies of software libraries for MapServer. When a user request reaches the MapServer binary through the Common Gateway Interface (CGI), the underlying MapServer code library, libmapserver2, uses a range of 16 linked libraries to handle different data formats, e.g., GIF, PNG, SVG, XML, processing tasks, e.g., PROJ, GDAL, PostgreSQL, but also map rendering tools, e.g., CAIRO, and typesetting, like FreeType and HarfBuzz. A significant number of additional dependencies exist in lower branches in the dependency tree, which are not included in the figure. Libraries that are jointly used by MapServer and GRASS GIS [36] emphasizing the occurring reuse of geospatial libraries among open source projects are *shaded*. * Geospatial libraries covered in this chapter

The origins of PROJ date back to the 1970s [23]. The United States Geological Survey (USGS) launched the development of two programs, one to plot maps [157] and PROJ to convert geodetic coordinates into several different cartographic projections [41]. PROJ evolved out of the earlier General Cartographic Transformation package (GCTp) Fortran code. The first release of *proj* was developed by Gerald Evenden in the early 1980s as a Ratfor program [158]. The second release in 1985 was ported to in C to run on Unix systems. The third release of *proj* from 1990, named PROJ.3, was expanded to support approximately 70 cartographic projections. A fourth release occurred in 1994, named PROJ.4, which is the current name of the software. The current version is PROJ v8.2.1 [155], which provides a new API and additional significant features [24].

The PROJ project comprises the following software tools:

The software library *libproj* as the OSGeo-foundational component of PROJ, which is used through its application programming interface (API) as a middleware library by many other open source projects. In addition, the applications *proj* and *ctt* are provided as command line interfaces (CLI) to libproj. The program *cs2cs* translates any pairs of coordinate systems, including support for datum translation. Geodesic (great circle) computations are implanted through the *geod* program. The Geospatial Integrity Investigation Environment (*gie* program) can be used for regression tests.

Within PROJ, reference information about geodetic objects, coordinate reference systems and their metadata, and coordinate operations between those CRS is internally in a unified database (proj.db). It contains data from the IOGP EPSG dataset, the IGNF (French national mapping agency) geodetic registry, and the Esri projection engine database.

The program *projinfo* enables queries about geodetic objects of in the database, import and export geodetic objects from/into WKT and PROJ strings, and display coordinate operations available between two coordinate reference systems (CRS).

Apart from conversions between and datum shifts within one CRS, including both cartographic projections and geodetic transformations, PROJ can handle dynamic reference frames and accurate coordinate transformations as needed with high accuracy GNSS. The terminology used within PROJ adheres to OGC and ISO-19100 geospatial standards.

PROJ implements abstract models according to the ISO-19111:2019 "Referencing by coordinates" standard [159], for geodetic reference frames (datums), coordinate reference systems, and coordinate operations. Geodetic objects can be imported and exported from and into the OGC Well-Known Text format (WKT) as Esri WKT, GDAL WKT 1, and WKT2:2019 (ISO 19162:2019 [160]).

30.7.3 GDAL/OGR

The Geospatial Data Abstraction Library (GDAL) is an open source C++ library for reading, writing, and transforming geospatial data. GDAL currently supports over 200 formats of raster and vector data [161, 162]. It is a translator library for working with raster data formats with an integrated component library, OGR, which provides support for vector formats. GDAL/OGR is the most widely used open source geospatial package, with distributions made available for Mac OS, Microsoft Windows, and Unix/Linux [163]. GDAL/OGR provides read/write and geoprocessing services for a long list of open source and proprietary GIS and mapping technologies, including MapServer, GRASS, QGIS, ArcGIS, OpenEV, OSSIM, FME, and Google Earth. API access is provided for C, C++, Python, Ruby, Java, and PERL languages. GDAL/OGR utilities can easily process large collections of data as well as large file sizes (> 4 GB) [164]. It also includes a set of useful command line utilities for working with data outside of a particular application.

GDAL relies on one common data model for all raster formats and one common model for all vector formats [165]. These abstract models permit the interoperability of datasets between platforms and services. Data models define the type and structure of information stored for a particular dataset, including semantics that express size, spatial reference, auxiliary metadata (raster), and feature definitions (vector).

GDAL/OGR includes utilities for reprojection, warping, compression, and format conversion for all data types. To allow for more efficient raster data storage and delivery, GDAL provides mosaicking and tiling capabilities. Other raster utilities support pixel-level image manipulation, rescaling, adding ground control points, and setting of nodata values. Utilities in OGR for working with vector data include dataset merging and tileindex creation. OGR also allows for the creation, manipulation, and subsetting of data attributes [17].

30.7.4 Java Topology Suite (JTS)

The Java Topology Suite is an open source Java library that provides an object model for planar geometry together with a set of fundamental geometric functions. The project was conceived and initiated by Mark Sondheim of the Geographic Data British Columbia (GDBC) branch of the BC Ministry of Environment, Land, and Parks. Its first implementation was started by Martin Davis and Jonathan Aquino, both working for Vivid Solutions at the time, in 2000, backed by funding obtained from GeoConnections and the Government of British Columbia. The first stable release was published in February 2002 [166].

Apart from the main requirement to create a Java API that would conform to the Simple Features Specification for SQL published by the Open Geospatial Consortium (OGC), the developers pursued the objective of producing a library that would be very robust and highly performant, as well as written in 100% pure Java.

The long history and its robustness have made the JTS library a core component for most Java-based open source geospatial applications [17]. A few examples of applications, libraries, and tools that make use of JTS are:

- The desktop GIS gvSIG [167], uDig [168], and OpenJump [169]
- The open source mapping server GeoServer [170]
- The open source Java geospatial library geotools [171]
- The spatial database H2GIS [172]
- The mobile Android app Geopaparazzi [173].

A subset of the library was ported to C/C++ under the library named GEOS, through which JTS reaches even more applications and tools. A small list of such projects is given by the spatial database PostGIS [174], the desktop GIS QGIS [135], and the data abstraction library GDAL [175]. Also ports in .NET and JavaScript are available under the NetTopologySuite and JSTS (https://openhub.net/p/jsts) projects.

In 2016/2017, Vivid Solutions agreed to bring the JTS project to the LocationTech [176] working group of the Eclipse Foundation. The transition brought three major changes. The namespace of the source code was changed from *com.vividsolutions* to *org.locationtech.jts*, and the license changed from the LGPL license to a new dual licensing system. JTS is currently licensed under the Eclipse Public License 1.0 and the Eclipse Distribution License 1.0, which is a BSD license. The codebase has been organized into modules for better clarity. The module structure of the version at the time of writing (1.16) is the following:

- jts-core: main implementation of functions and algorithms
- jts-tests: extensive test suite (Fig. 30.19)
- jts-io: read and write tools for geometries
- jts-example: usage examples of the library
- jts-lab: experimental playground with unstable functionalities
- jts-app: test builder application for defining tests.

The JTS project supplies a tool named testbuilder, through which the JTS library can be used in standalone mode as an application and allows the user to browse and test all its features. The application allows the user to define geometries in the dedicated panels at the bottom of the main view in WKT [177], WKB [178], or GML [179] format or to draw them in the coordinates view using simple drawing tools. The left-hand panel features a *Geometry Functions* tab that reveals the list of available JTS functions. A nonexhaustive list of features is given by:

- Spatial predicates based on the DE-9IM model
- Overlay functions (as for example intersection, difference, union)
- Buffer computation including single side buffer and cap and join modification
- Convex hull
- Geometric simplification (Douglas-Peucker and Visvalingam-Whyatt) and densification
- Delaunay triangulation
- Voronoi diagram

- Polygonization
- Linear referencing
- Topological validity checking
- Metric functions for the calculation of distance between geometries and area and length.

Moreover, the library features robust line segment intersection and line-point orientation, as well as efficient point-in polygon testing. It also implements a very efficient cascading union approach for the union of polygon collections. JTS also contains a set of spatial index structures including STR-tree (the fastest, cannot be modified once built), Quadtree (slower, but supports insert and delete), KD-tree, Interval R-tree, and Monotone Chains.

30.7.5 GeoTools

The GeoTools [180] project is a free and open source library whose objective is to provide a toolkit for the development of standards compliant geospatial solutions. The project was initiated as part of the master project of James Macgill under the tutorship of Ian Turton at the University of Leeds in 1996. Its first objective was to provide a toolkit that would allow the visualization of maps in the web browser through a new and very promising technology that had been recently released: the Java applet [181].

The first version of the GeoTools project was not designed with a scalable architecture and particular standards in mind. The growing interest around the library pushed the development rather quickly, and with time, the limits of the first architecture came to the surface, calling for a refactoring of the project.

With GeoTools 2, the development force grew from the two original developers to an open source community of developers. For the second version, the codebase was basically redesigned to take full advantage of the Java platform and to allow modularization. In 2002, code from the SEAGIS [182] project was brought into GeoTools, providing coordinate transformation services, grid coverage, and rendering implementations.

Today, the GeoTools library is a large open source Java project that counts a number of active contributors and a rich feature portfolio. It is the de-facto open source Java library for the creation of geospatial applications and used by projects such as, for example, GeoServer [170], uDig [168], the Hortonmachine [183], Geomajas [184], and Geospatial ETL Talend [185]. It implements the following standards:

- ISO 19109 General feature model and ISO 19125-1 Simple features
- OGC Grid Coverage representation of raster information
- OGC Styled Layer Descriptor
- OGC Filter and Common Constraint Language (CQL)
- Clients for Web Feature Service, Web Map Service, Web Coverage Service, and Web Process Service
- ISO 19107 Spatial schema (geometry).

As such it provides a robust toolkit for the creation of geospatial applications that need standards compliance. Its feature richness assures that these applications have support for a number of raster and vector data formats, being it online or file based (examples are GeoTIFF, grassraster, GTOPO30, JP2K, PostGIS, spatialite, shapefile, GeoJSON, and arcgrid). GeoTools also leverages the ImageIO-EXT [186] library, which allows access to several raster formats from the GDAL library [9, 175].

To better understand the structure of the GeoTools project, it can be split into levels. The interface level contains the API part, the standards part, which finds its space in the OpenGIS modules, and the JTS part, which is one of the base libraries on which GeoTools is built. On top of the interface level, the implementation level supplies, beyond others, the data model, the metadata, and referencing parts, as well as the rendering and CQL functionalities. On top of the implementation level, many community contributed plugins and extensions provide a large feature set. That way, GeoTools provides data format support for files and online standard services (for example, WMS, WCS, WFS) including spatial databases. In the list of functionalities provided through modules, one can find graph and networking support (shortest path), color brewer support [187], and plugins to handle common and less common data formats as GeoJSON, geobuf, MBTiles, GeoPackage, and dxf.

The GeoTools library comes with a swing module, which can be used to create desktop applications featuring a graphical user interface (GUI). The module also contains a couple of small example applications that showcase the usage of the library providing a window with a map view and some tools to interact with the map through pan and zoom, do some simple map styling, and load file based data formats as shapefile vector files and GeoTIFF raster datasets (Fig. 30.20).

At the time of writing this book chapter, GeoTools has reached version 26 and is licensed under LGPL to foster its use on commercial and proprietary projects. GeoTools graduated to be an OSGeo project in July 2008 and has since then been part of the OSGeo branded geospatial libraries together with OSSIM, Orfeo Toolbox, GDAL/OGR, and GEOS.

30.7.6 Leaflet

Leaflet [18] is an open source, JavaScript library built for interactive maps for the web. Originally created by Vladimir Agafonkin, Leaflet is designed to be lightweight, extensible, and easy to use for various purposes. The software has been

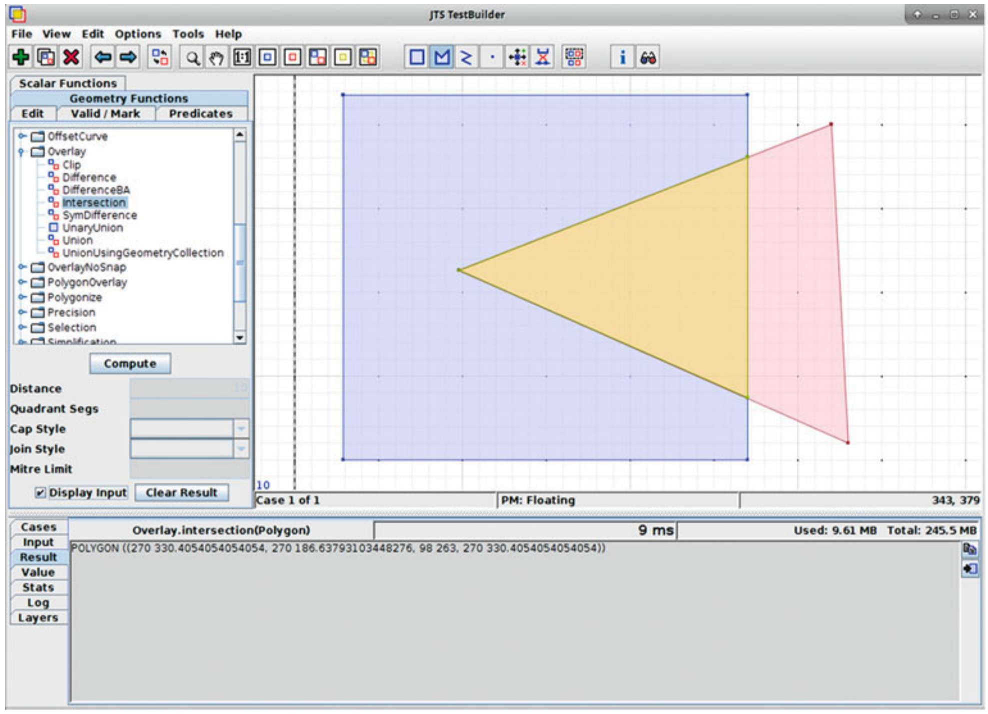


Fig. 30.19 The testbuilder tool. The main view shows the panel of the available geometry functions and the result of an intersection applied to the two polygons supplied as input WKT geometries

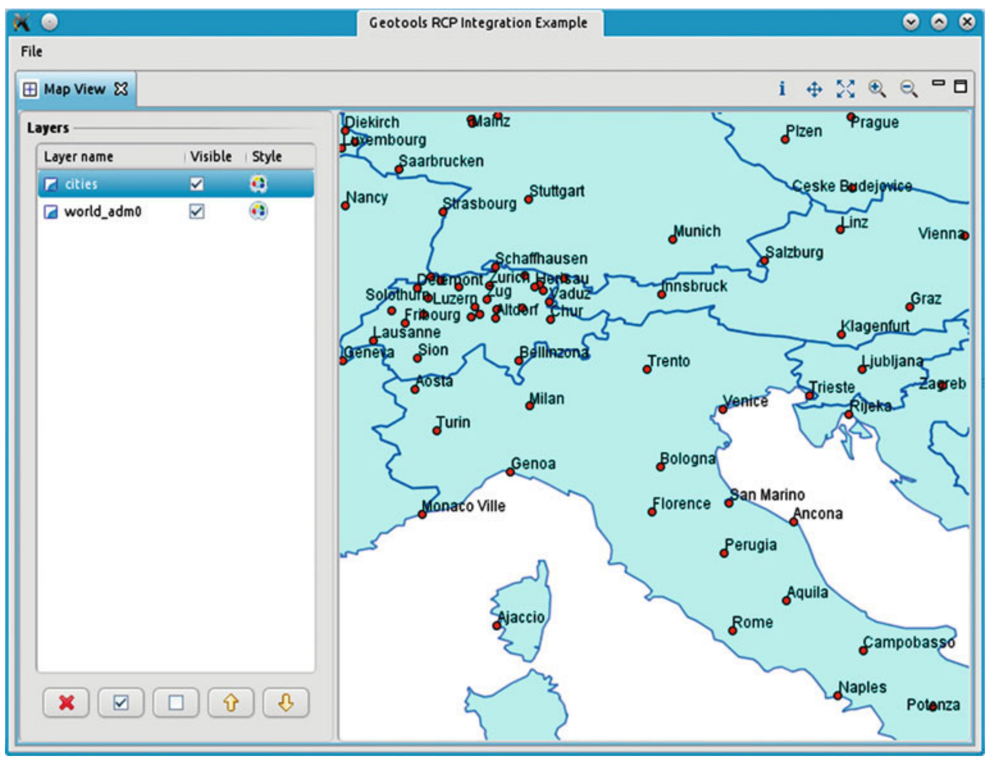


Fig. 30.20 An example of the integration of the SWT GUI module inside an RCP application [188]. The same way it is possible to create a Swing GUI through the Swing module. The example shows the maps of the countries and cities of central Europe

widely adopted by a diverse set of technology companies, news organizations, and others building interactive maps.

Agafonkin began privately working on a secret web map library in 2008 in response to his desire for a simpler library to use on a work project [189]. Popular libraries at the time were either too large or had APIs that did not meet the project's expectations. The library developed by Agafonkin eventually became Leaflet. As the project creator states, Leaflet was "born as a protest against bloat, clutter, and complexity" [190].

Core to Leaflet's original design goals were support for modern and legacy browsers, usability, and performance [191]. These goals have held true throughout the project's history, helping to keep the library well focused and simple, which has aided its popularity. An important result of its emphasis on usability is a well-documented codebase with examples and tutorials [192].

Another key design decision that guided the development of Leaflet was to build the library with extensibility in mind. This decision allowed for the library's core functionality to remain lightweight and useful in a variety of contexts, while also not limiting additional functionality through plugins. The software's extensibility has also allowed the community surrounding Leaflet to flourish, giving developers a venue for making valuable contributions and users a multitude of options to customize or enhance their maps.

Leaflet offers core functionality of adding common geospatial data types to an interactive web map application. Some of the supported standards-based data formats include Open Geospatial Consortium (OGC) Tile Map Service (TMS) [193], Web Map Service (WMS) [194], and GeoJSON [195]. The library also enables overlaying static images or video and has extensive support for vector overlays.

Leaflet offers an extensible, object-oriented programming [196] (OOP) approach to building interactive digital map applications. This framework allows adopters to extend the library's functionality in new and interesting ways and to share this work through Leaflet's plugin ecosystem and registry. The registry currently lists over 400 plugins, which represent a broad set of domains, needs, and interests. With little overhead for users, these plugins add support for common or novel visualization capabilities like heatmaps, extend compatibility with new data types like vector tiles or IIIF [197], and offer countless ways to interact with and imagine maps.

30.8 Virtual Globes

Virtual globes allow us to interactively display and query GIS data sources through a rendering of a planetary globe. The elevation and angle of the position of the observer can be dynamically changed in real time. Virtual globes allow us to dynamically visualize data in 2-D, 2.5-D, 3-D, and over time. While many proprietary virtual globes (e.g., Google Earth) can only be accessed online based on remote map servers, this chapter showcases two open source solutions, which can also be used locally or in a client–server setting.

30.8.1 NASA WorldWind

NASA WorldWind (NWW) is an open source virtual globe project that was started in 2002 by *NASA Learning Technologies*, a program to produce educational material from and for NASA datasets. It was one of the first NASA programs to be released as open source under permissive license in order to widespread its implementation in both open and proprietary environments and be an interface that would foster greater participation.

NWW was developed in the same timeframe as the virtual globe of the company Keyhole, which was developing what would later be Google Earth, so technically speaking, the project is older than the better known project Google Earth.

One of NWW's objectives was to be able to load and investigate datasets from NASA. The first version of NWW, developed with .NET technology, provided a standalone application with an extensive suite of plugins. Apart from the Earth, it also featured data for other worlds, such as the Moon, Mars, Venus, and Jupiter.

Being limited to the Windows operating system did not fully allow for larger accessibility and implementation by NASA and a world of others. Hence, a new version was developed using Java. So, in the summer of 2005, this gave birth to the WorldWind Java project as an SDK (Standard Development Kit). WorldWind Java allowed developers to create their own applications based on it, greatly facilitating government agencies and private enterprise to develop entirely custom applications tailored to their specific needs [198].

The project has been very successful since its conception, and at the time of writing, three versions, sharing a similar API (and, therefore, being very attractive to developers), are growing in parallel:

- WorldWind Java: the oldest project, a Java SDK for Desktop applications
- Web WorldWind: a virtual globe SDK for web applications
- WorldWind Android: an SDK to use the virtual globe in Android applications.

Starting from the main homepage of the project [199] it is possible to access all three project spaces, all of which provide extensive documentation, tutorials, and examples to help new developers to quickstart into development of applications with the WorldWind ecosystem.

WorldWind Java (WWJ)

WWJ, but in general all projects of the WorldWind family, allows us to create on top of the virtual globe point, lines, and polygon layers. It is also possible to load 3-D models into its layers, as for example, the ISO standard format COLLADA. Imagery can be placed in layers on the globe through standard Open Geospatial Consortium (OGC) WMS services or by loading local raster files as GeoTIFFs. The default elevation model, which makes use of the SRTM [200] datasets, can be changed with a more precise set of elevation data that is usually available from local government agencies

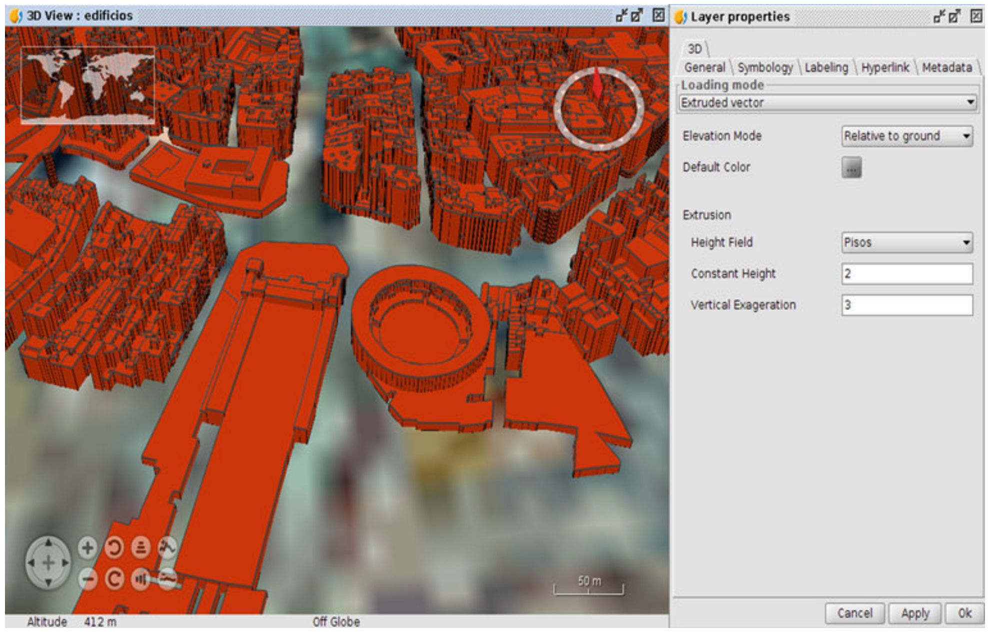


Fig. 30.21 WWJ integrated in GIS gvSIG. A view of the extruded buildings of Valencia, Spain

(Fig. 30.21). It is possible to automate navigation around the globe through the control of the camera movement, providing the user with a fly-to feeling.

An open source GIS that implements all of the abovementioned features and can almost be seen as a showcase of the WWJ features integrated in a GIS is gvSIG [16].

WWJ is available under the NASA Open Source Agreement version 1.3 [201].

WorldWind Web (WWWeb)

In December 2017, the first release of WWWeb was published, version 0.9.0. The web version of NWW gained support from a wider community, and an important contributor joined the project, the European Space Agency (ESA). The application is written in HTML5 and JavaScript and released under Apache License, version 2.0. WWWeb is distributed bundled with GeoServer [170], serving as a server-side mapping engine. Similar to the Java version, the web version features a collection of geometric shapes, a large number of built-in high-resolution datasets, and has the ability to display imagery, terrain, and geographic data from standard OGC services as WMS (Fig. 30.22).

WorldWind Android (WWA)

In the current digital mobile age, a mobile version of the WorldWind virtual globe could not be missing, and in 2016, the first public version 0.1.0 of WWA was released. It was a prototype release that featured a basic navigable 3-D globe and was capable of displaying data from WMS services. Since then many features have been added and at the time of writing, while the project has reached version 0.8.0, it supports various geometry shapes, overlay of OGC services, picking of features on the map, as well as globe navigation. WWA is available under the NASA Open Source Agreement version 1.3 [201].

NWW Project Outage

In May 2019, the WorldWind project at NASA was suspended due to missing funds [202]. It is important to highlight that the web version is also supported by the European Space Agency (ESA). Due to its open source nature, the WorldWind project started to reorganize itself and created community forks to advance with the development. For the NWW Java and Web projects community, forks got into active development (JAVA [203], WEB [204]). While the Android version also seemed to have a fork on the GitHub site [205], it was not clear whether a critical mass of developers would be able to create a large enough community to push the project forward. By the end of the same year NASA announced a renewed commitment to WorldWind [206], which resulted into a new release of both the Web and Java WorldWind projects in August 2020. No Android version has been released yet at the time of writing.

30.8.2 GPlates

The open source GPlates software [207, 208] combines spatiotemporal data fusion, analysis, and visualization through geological time. GPlates provides the unique capability to reconstruct geodata attached to tectonic plates to develop mathematical models that describe how the plates and the boundaries between them have evolved through deep time. The software allows users to explore the tectonic evolution of the Earth through cycles of supercontinent assembly, fragmentation, and dispersal. GPlates applications include

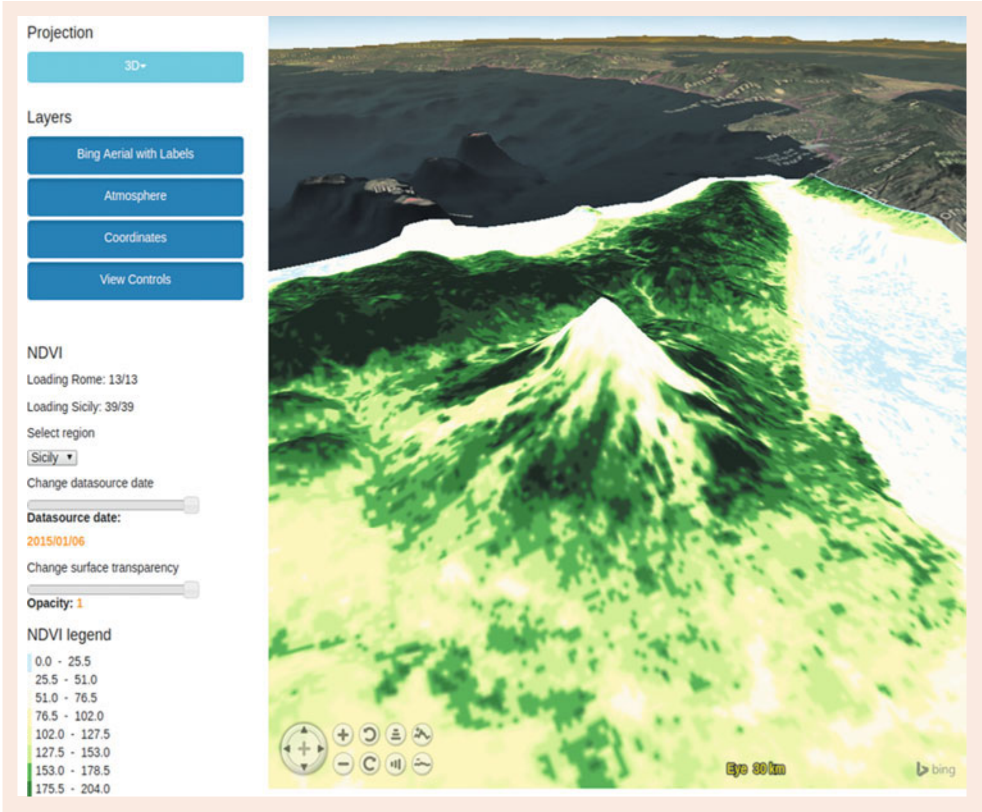


Fig. 30.22 An example application of WWWeb showing a map of NDVI overlaid on BING imagery of Sicily, Italy. The elevation exaggeration has been increased to better appreciate the 3-D view. The image was created using the NDVI Viewer demo, which can be found on the WWWeb examples page (https://worldwind.arc.nasa.gov/web/examples/)

tectonics, geodynamics, basin evolution, orogenesis, deep Earth resource exploration, paleobiology, paleoceanography, and paleoclimate.

The GPlates Geological Information Model (GPGIM) represents a formal specification of geological and geophysical data in a time-varying plate tectonics context [209]. It provides a framework in which relevant types of geological data are attached to a common plate tectonic reference frame, allowing the data to be reconstructed in a time-dependent spatiotemporal framework. The GPlates Markup Language (GPML), being an extension of the open standard Geography Markup Language (GML) [210], is both the modeling language for the GPGIM and an XML-based data format for the interoperable storage and exchange of data modeled by it. The GPlates software implements the GPGIM, allowing researchers to query, visualize, reconstruct, and analyze a rich set of geological data including numerical raster data. The GPGIM has been extended to support time-dependent georeferenced numerical raster data by wrapping GML primitives into the time-dependent framework of the GPGIM. GPlates is interoperable with other GIS tools, including QGIS, ArcGIS, GeoMapApp, and others (i.e., it can read and write shapefiles), with reconstructed GPlates data that can be exported to be processed and plotted seamlessly using the open source Generic Mapping Tools [211].

In GPlates, users can build regional or global plate motion models, import their own data, and digitize features. GPlates can handle paleomagnetic data, create and display virtual paleomagnetic poles, and derive absolute plate rotations from them. GPlates allows users to interactively investigate alternative fits of the continents, to test hypotheses of supercontinent formation and breakup through time, and to unravel the evolution of tectonically complex areas such as the Caribbean and Southeast Asia. Numerical and color raster files and images in a variety of formats can be loaded, assigned to tectonic plates, age-coded, and reconstructed through geological time. The software also allows the exporting of image sequences for animations or for publication-quality figure generation as vector graphics files. Plates and plate boundaries through time can be visualized over time-dependent rasters, including seismic mantle tomography image stacks.

All features in GPlates store their geometry in present-day coordinates, so that different rotation models can be used to reconstruct them backward in time. GPlates can reconstruct a variety of present-day vector geometries, including points, multipoints, polylines, and polygons. Although these geometries are reconstructed using Euler rotations, their overall shape remains the same, and, hence, they are often referred to as *rigid* or *static* geometries. When a new feature is digitized at a past geological time, the digitization tool reverse reconstructs the digitized geometry forward in time to the present day. This way, when the newly created feature is reconstructed back to its digitization time, it will match the

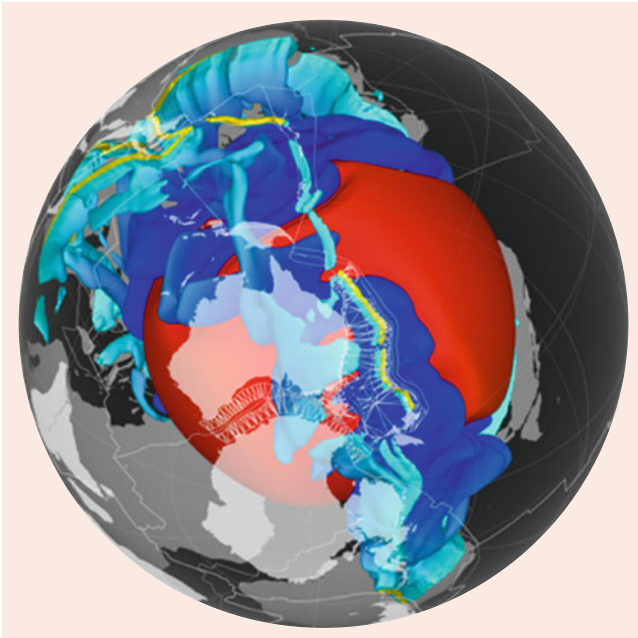


Fig. 30.23 A GPlates visualization of modeled mantle structure showing Australia and surrounding continents (*translucent gray*) 80 million years ago, with plate boundaries shown as *thin white lines*, active lithospheric extension indicated by meshes and mantle temperature illustrated by cold, sinking slab material colored by depth (*yellow* and *magenta* in the upper mantle and *blue* in the lower mantle) and large mantle upwellings rising from the core-mantle boundary (*red*)

location of the original digitized geometry. These reconstructions use information stored inside a feature (in the form of feature properties) to determine how to reconstruct the feature. Most features are simply reconstructed using a single-plate identification number (Plate ID) that looks up a rotation model. Other features have more complex reconstructions, such as a mid-ocean ridge, that use two Plate IDs to calculate spreading between the neighboring plates. In addition to reconstructing regular geometries (points, lines, and polygons), GPlates supports interactive reconstruction of very large raster datasets [212]. Raster data can be imported from a variety of color image formats containing Red Green Blue Alpha (RGBA) color data, as well as numerical image formats containing floating-point and integer data (including NetCDF and GeoTIFF). Numerical raster data are visualized by selecting a color palette that is either built-in or loaded from a GMT [211] regular CPT (color palette table) file.

One of the fundamental innovations in GPlates is the design to enable the linking of plate tectonic models with mantle convection models [213]. The software allows the construction of time-dependent plate boundary topologies, as well as exporting plate polygons and velocity time sequences. Numerical mantle convection model output, e.g., in the form of mantle temperature volumes, can be imported with plate tectonic reconstructions overlain (Fig. 30.23 [207]). This is enabled by support for the interactive visualization of subsurface 3-D scalar fields. Scalar fields can be visualized either as isosurfaces or cross sections. An isosurface represents the surface through the scalar field with a specific scalar value known as the isovalue. Isosurfaces are rendered by tracing a ray through the scalar field, at each screen pixel in the globe view, until the isosurface is found. To achieve this, a ray-tracing program is executed on the graphics hardware for each pixel. This approach is particularly suited to modern graphics hardware because the rays (pixels) are independent of each other and, hence, can be distributed in parallel across the thousands of computational units in the graphics processing unit (GPU). This enables real-time interactivity when the isosurface is rerendered each time the view changes or the user adjusts the isovalue.

GPlates deforming plate functionality allows users to define the spatial and temporal extent of diffuse plate boundaries and model the deformation of these regions through time [214]. These are regions combining extension, compression, and shearing that accommodate the relative motion between rigid blocks that follow the traditional concept of plate tectonics. Users can explore how strain rates, stretching/shortening factors, and crustal thickness evolve through space and time within deforming regions, and interactively update the kinematics associated with deformation to investigate how these parameters are influenced by alternative plate reconstruction scenarios. The geometries that define regions of deformation change over time in response to the user-defined kinematics, and the consequences of these changes can be quantified and represented using stretching/shortening factors. Together, the new tools form the basis for building reconstructions that quantitatively describe the cycles of rifting, mountain building, and intracontinental shearing that accompany supercontinent assembly and dispersal.

30.9 OSGeoLive

The OSGeoLive project [21, 215] provides collections of fully-operational versions of popular free geospatial software as self-contained Lubuntu-based bootable Linux distribution for geospatial applications. The collections are published as bootable ISO images to be used on a USB thumb drive or DVD, and a pre-made virtual machine with additional tools and data to be used in virtual machine applications such as VirtualBox, VMWare, or KVM.

The OSGeo project started in 2008. The first OSGeoLive release was published at the FOSS4G 2008 conference. The current release is OSGeoLive 14.0 (Fig. 30.24) [216]. OSGeoLive images can also be installed on a local computer.

Through its website the project provides OpenHUB-based metrics for all geospatial software hosted within the current collection [217], which are derived from the projects' code repositories.

Fig. 30.24 Overview of the user menus for geospatial FOSS tools in the OSGeoLive 14.0 desktop environment, comprising varieties of geospatial browser clients, databases, desktop GIS, navigation and maps, spatial tools, and map services

The contents of a OSGeoLive distribution are grouped in the sections desktop GIS, browser facing GIS, web services, data stores, navigation and maps, spatial tools, domain-specific GIS, data, geospatial libraries, and geospatial standards. In addition to the preinstalled geospatial software, documentation, sample datasets and a comprehensive presentation are also included [218].

As the collection of geospatial software tools is renewed annually, some tools that were part of the collection in former years are retired. Earlier releases of OSGeoLive are available from the OSGeoLive Documentation Archive and provide access addtional older software tools and versions [219]. The OSGeoLive website provides a listing of the contents of all previous OSGeoLive releases, which can be used to identify and access earlier releases if a particular software is needed (https://live.osgeo.org/en/prior_applications.html).

The system requirements of OSGeoLive are lightweight to ensure that the system will also run on older hardware. A hard drive is not required.

The current OSGeo collection comprises the following geospatial software tools and multiple sample datasets. Software discussed in the chapter is marked with *:

- **Desktop GIS**: GRASS GIS 7*, gvSIG desktop*, OpenJUMP, QGIS*, SAGA GIS, uDig
- **Browser Facing GIS Clients**: OpenLayers, Leaflet*, Cesium, Geomajas, Mapbender, GeoMoose, GeoNode, GeoExt, Cesium (a virtual globe)
- **Data Stores**: PostGIS*, SpatiaLite, rasdaman*, pgRouting
- **Navigation and Maps:** Open Street Map JOSM, GpsPrune, Marble (a virtual globe), OpenCPN, zyGrib
- **Spatial Tools:** Jupyter Notebook, GMT, Mapnik, MapSlicer, Monteverdi, Orfeo ToolBox, R Statistics
- **Geospatial Libraries:** GDAL*, GeoTools*, GEOS, PROJ*, libLAS, JTS*
- **Geospatial Webservices:** 52North WxS and SWE services, deegree, GeoNetwork, GeoServer*, mapProxy, ncWMS, EOxServer, istSOS, MapCache*, MapServer*, pycsw, PyWPS, QGIS Server, Actinia, T-REX, and Zoo

References

1. Neteler, M.: actinia: geoprocessing in the cloud, https://av.tib.eu/media/47043, (2020). https://doi.org/10.5446/47043
2. Reed, J.: Fixing GIS data discovery, https://av.tib.eu/media/31700, (2014). https://doi.org/10.5446/31700
3. Antonello, A. (HydroloGIS), Franceschi, S. (HydroloGIS-Free University Of Bolzano): Digital field mapping with Geopaparazzi and gvSIG, https://av.tib.eu/media/20316, (2016). https://doi.org/10.5446/20316
4. Antonello, A., Franceschi, S.: Geopaparazzi: state of the art of the digital field mapping application, https://av.tib.eu/media/43378, (2019). https://doi.org/10.5446/43378
5. Tzotsos, A., Kralidis, T.: pycsw project status 2019, https://av.tib.eu/media/43465, (2019). https://doi.org/10.5446/43465
6. Čepický, J.: State of PyWPS, https://av.tib.eu/media/43436, (2019). https://doi.org/10.5446/43436
7. Kralidis, T., Broecke, J.V.: GeoHealthCheck: A Quality of Service Monitor for Geospatial Web Services, https://av.tib.eu/media/43344, (2019). https://doi.org/10.5446/43344
8. Garnett, J.: GeoServer Orientation, https://av.tib.eu/media/14076, (2020). https://doi.org/10.5446/14076
9. Aime, A.: Raster Data In GeoServer And GeoTools: Achievements, Issues And Future Developments, https://av.tib.eu/media/32103, (2015). https://doi.org/10.5446/32103
10. Wessel, P.: 20 years with GMT – The generic Mapping Tools, https://av.tib.eu/media/19869, (2011). https://doi.org/10.5446/19869
11. Neteler, M.: GRASS GIS 6.4 development visualization from 1999 to 2013, https://av.tib.eu/media/14652, (2013). https://doi.org/10.5446/14652
12. Metz, M., Neteler, M., Landa, M., Andreo, V., Petrasova, A., Delucchi, L., Lennert, M., GRASS Development Team: State of GRASS GIS Project: 35 years is nothing!, https://av.tib.eu/media/43513, (2019). https://doi.org/10.5446/43513
13. Neteler, M.: GRASS GIS 6.4 development visualization from 1999 to 2013, https://av.tib.eu/media/14652, (2013). https://doi.org/10.5446/14652
14. Carrera, M.: gvSIG Suite: a complete catalog of open source software solutions, https://av.tib.eu/media/43948, (2018). https://doi.org/10.5446/43948
15. Martinez, O.: gvSIG Desktop: written in Java, extendable through Scripting, https://av.tib.eu/media/43972, (2018). https://doi.org/10.5446/43972
16. Olmos, Ó.M., Izquierdo, C.M.: 3D Tools in gvSIG using NASA World Wind, https://av.tib.eu/media/20298, (2016). https://doi.org/10.5446/20298
17. Hughes, J., Emanuele, R., Garnett, J., Davis, M.: State of JTS, https://av.tib.eu/media/41095, (2017). https://doi.org/10.5446/41095
18. Agafonkin, V.: Leaflet: Past, Present, Future, https://av.tib.eu/media/15548, (2013). https://doi.org/10.5446/15548

19. Girvin, S., Morissette, D., Smith, M.: State of MapServer 2019, https://av.tib.eu/media/43440, (2019). https://doi.org/10.5446/43440
20. Bonfort, T.: MapCache: Fast and Featureful tile serving from the MapServer project, https://av.tib.eu/media/32099, (2015). https://doi.org/10.5446/32099
21. Emde, A.: OSGeoLive: Your Open Source Geospatial Toolkit, https://av.tib.eu/media/52538, (2021). https://doi.org/10.5446/52538
22. Kastl, D., Vergara, V.: Shortest Path search in your Database and more with pgRouting, https://av.tib.eu/media/20359, (2016). https://doi.org/10.5446/20359
23. Butler, H.: A history, status report, and outlook of Proj.4, https://av.tib.eu/media/40916, (2017). https://doi.org/10.5446/40916
24. Evers, K.: PROJ – A status report and look into the future, https://av.tib.eu/media/43965, (2018). https://doi.org/10.5446/43965
25. Bernasocchi, M.: QGIS is dead, long live QGIS! – The very best new features of QGIS 3.x, https://av.tib.eu/media/43397, (2019) https://doi.org/10.5446/43397
26. Baumann, P.: Datacubes on Steroids with ISO Array SQL Open Source, Open Standards, Open Terabytes, https://av.tib.eu/media/41961, (2018). https://doi.org/10.5446/41961
27. Tzotzos, A.: Open Source Geospatial Foundation (OSGeo), https://av.tib.eu/media/55248 (2021). https://doi.org/10.5446/55248
28. Raghavan, V.: Sol Katz Award 2016 & Students Award, https://av.tib.eu/media/20468, (2016). https://doi.org/10.5446/20468#t=16:50,24:07.
29. Crăciunescu, V.: Awards ceremony, https://av.tib.eu/media/43542, (2019). https://doi.org/10.5446/43542#t=20:28,26:07
30. Wikipedia contributors: Map Overlay and Statistical System (MOSS) (2019). https://en.wikipedia.org/wiki/Map_Overlay_and_Statistical_System, Accessed 2 May 2022
31. Reed, C.: MOSS – A historical perspective (2004). https://de.scribd.com/document/4606038/2004-Article-by-Carl-Reed-MOSS-A-Historical-perspective, Accessed 2 May 2022
32. Reed, C.N.III., Katz, S., Frosh, R., Davidson, J., Taylor, A., Lee, J.: OSGeo/MOSS: MOSS Archive 1.2.3. Zenodo (2022). https://doi.org/10.5281/ZENODO.5140318
33. https://github.com/OSGeo/MOSS, Accessed 18 Aug 2021
34. Löwe, P., Nartišs, M.: MOSS: Back to the future, https://doi.org/10.5281/zenodo.5761252, Accessed 6 Jan 2022 (2021)
35. https://www.osgeo.org/projects/moss/, Accessed 18 Aug 2021
36. Landa, M., Neteler, M., Metz, M., Petrasova, A., Glynnc, HamishB, Vaclav Petras, Huhabla, Huidae Cho, Delucchi, L., Pietro, Barton, M., Zigo, T., Chemin, Y., Māris Nartišs, Ostepok, Kudrnovsky, H., Mlennert, Larsson, N., Blumentrath, S., Kyngesburye, W., Kladivova, L., Aghisla, Andreo, V., 積丹尼 Dan Jacobson, Ovsienko, D., Madi, Helena Mitasova, Tawalika, C., Caitlin, H.: OSGeo/grass: GRASS GIS 7.8.7RC1. Zenodo (2022). https://doi.org/10.5281/ZENODO.5176030
37. Inman, R.J.: GRASS, https://av.tib.eu/media/12963, (1987). https://doi.org/10.5446/12963
38. GRASS Development Team: Historical notes (1998). https://grass.osgeo.org/home/history, Accessed 2 May 2022
39. Evenden, G.I.: Forward and inverse cartographic projection procedures. Open-File Report 83-625. U.S. Geological Survey (1983). p. 76 https://pubs.er.usgs.gov/publication/ofr83625, Accessed 2 May 2022
40. Evenden, G.I.: Cartographic projection procedures for the unix environment – A user's manual (1995). https://pubs.usgs.gov/of/1990/of90-284/ofr90-284.pdf, Accessed 2 May 2022
41. GRASS Development Team: GRASS mailing list community evolution (2016). wiki.osgeo.org/wiki/Annual_General_Meeting_2020, Accessed 6 Jan 2022
42. MapServer Development Team: MapServer conferences (2012). Internet Archive. https://web.archive.org/web/20220120030437/https://trac.osgeo.org/mapserver/wiki/Conferences, Accessed 2 May 2022
43. OSGeo Foundation: Projects (2019). https://www.osgeo.org/projects, Accessed 2 May 2022
44. Löwe, P., Neteler, M. Goebel, J. Tullney, M.: Towards OSGeo best practices for scientific software citation: Integration options for persistent identifiers in OSGeo project repositories. (2017) Zenodo. https://doi.org/10.5281/zenodo.1034878, Accessed 6 Jan 2022
45. Sanz, J.: OSGeo charter members (2018). http://bl.ocks.org/jsanz/raw/779f9b9954b92461fa50, Accessed 2 May 2022
46. OSGeo Foundation: Annual general meeting (2018). https://wiki.osgeo.org/wiki/Annual_General_Meeting_2018, Accessed 2 May 2022
47. OSGeo Foundation: Events (2017). https://osgeo.org/history/events-history, Accessed Jan 6 2022
48. OSGeo: Sol Katz Award (2005). https://www.osgeo.org/community/awards, Accessed 24 Feb 2022
49. GNU Project: What is free software? (1996). https://www.gnu.org/philosophy/free-sw.html, Accessed 2 May 2022
50. Free Software Foundation: What is free software? (2019). https://fsfe.org/about/basics/freesoftware.en.html, Accessed 2 May 2022
51. Perens, B.: Debian's "social contract" with the free software community (1997). https://lists.debian.org/debian-announce/1997/msg00017.html, Accessed 2 May 2022
52. Apache Software Foundation: Apache licenses (2019). http://www.apache.org/licenses/, Accessed 2 May 2022
53. Linux Information Project: BSD license definition (2005). http://www.linfo.org/bsdlicense.html, Accessed 2 May 2022
54. Linux Information Project: MIT license definition (2009). http://www.linfo.org/mitlicense.html, Accessed 2 May 2022
55. Free Software: The history of the GPL (2001). https://www.free-soft.org/gpl_history/, Accessed 2 May 2022
56. GNU Operating System: GNI general public license 3 (2007). https://www.gnu.org/licenses/gpl-3.0.html, Accessed 2 May 2022
57. QGIS Project: QGIS user guide (2018). https://docs.qgis.org/2.18/pdf/en/QGIS-2.18-UserGuide-en.pdf, Accessed 2 May 2022
58. Qt Project: Homepage of the Qt project (2018). https://www.qt.io/, Accessed 2 May 2022
59. QGIS Project: QGIS GitHub repository (2018). https://github.com/qgis/QGIS, Accessed 2 May 2022
60. QGIS Contributors: QGIS. Zenodo. (2022). https://doi.org/10.5281/ZENODO.5869837
61. QGIS Project: QGIS Python plugins repository (2018). https://plugins.qgis.org, Accessed 2 May 2022
62. Westervelt, J.: GRASS roots. Proceedings of the FOSS/GRASS users conference, Bangkok (2004). https://oldgeni.isnew.info/grass/Jim-Westervelt.pdf, Accessed 2 May 2022
63. Neteler, M., Bowman, M.H., Landa, M., Metz, M.: GRASS GIS: A multi-purpose Open Source GIS. Environ. Model. Softw. **31**, 124–130 (2012)
64. Petrasova, A., Harmon, B., Tabrizian, P., Petras, V., Mitasova, H.: Tangible modeling with Open Source GIS. Springer, Cham, p. 138 (2018)
65. Jasiewicz, J., Stepinski, T.F.: Geomorphons – a pattern recognition approach to classification and mapping of landforms. Geomorphology **182**, 147–156 (2013)
66. Gebbert, S., Pebesma, E.: A temporal GIS for field based environmental modeling. Environ. Model. Softw. **53**, 1–12 (2014)
67. Rocchini, D., Neteler, M.: Let the four freedoms paradigm apply to ecology. Trends Ecol. Evol. **27**(6), 310–311 (2012)
68. Neteler, M., Mitasova, H.: Open source GIS: A GRASS GIS approach, 3rd edn. Springer, New York, p. 406 (2008)

69. Neteler, M., Gebbert, S., Tawalika, C., Bettge, A., Benelcadi, H., Löw, F., Adams, T., Paulsen, H.: Actinia: Cloud based geoprocessing. In: Soille, P., Loekken, S., Albani, S. (eds.) Proceedings of the 2019 Conference on Big Data from Space (BiDS'2019) EUR 29660 EN. pp. 41–44. Publications Office of the European Union, Luxembourg (2019) https://doi.org/10.5281/zenodo.2631917
70. Tawalika, C., Weinmann, A., Riembauer, G., Metz, M., Haas, J., Jansen, M., Herrera Maldonado, J.A., Gebbert, S., Neteler, M.: actinia-core. Zenodo. https://doi.org/10.5281/ZENODO.5865317 (2022)
71. Petras, V., Petrasova, A., Harmon, B., Mitasova, H.: Integrating free and Open Source solutions into Geospatial science education. ISPRS Int. J. Geoinf. **4**(2), 942–956 (2015)
72. GRASS GIS: Applications (2018). https://grass.osgeo.org/documentation/applications/, Accessed 2 May 2022
73. gvSIG Project: gvSIG homepage (2019). http://www.gvsig.com, Accessed 2 May 2022
74. Anguix, Á., Carrera, M., Cerro, J.D., Martínez, Ó., Higón, J.V., Peñarrubia, F.J., Martínez, C., Rodrigo, J., Díaz, F., Olivas, J., Brodín, I.: gvSIG. Zenodo (2022). https://doi.org/10.5281/ZENODO.5849091
75. gvSIG: Desktop project user guide (2017). http://downloads.gvsig.org/download/web/en/build/html/index.html, Accessed 2 May 2022
76. gvSIG: Scripting documentation for gvSIG desktop (2017). http://downloads.gvsig.org/download/web/en/build/html/scripting_devel_guide/2.3/index.html, Accessed 2 May 2022
77. HydrologGIS: Geopaparazzi (2017). https://geopaparazzi.github.io/geopaparazzi, Accessed 2 May 2022
78. GNU Operating System: GNU Affero general public license (2016). https://www.gnu.org/licenses/agpl-3.0.en.html, Accessed 2 May 2022
79. McKenna, J., PostGIS Team: PostGIS 3.2.0, The Olivier Courtin Release. Zenodo (2021). https://doi.org/10.5281/ZENODO.5879631
80. PostGIS: Home page (2022). https://postgis.net, Accessed 21. Feb 2022
81. Free Software Foundation: GNU general public license 3 (1991). http://www.gnu.org/copyleft/gpl.html, Accessed 2 May 2022
82. Open GIS Consortium: OpenGIS® simple features specification for SQL (revision 1.1) (1999). https://portal.opengeospatial.org/files/?artifact_id=829, Accessed 2 May 2022
83. International Organization for Standardization: ISO/IEC 13249-3: Information technology – Database languages – SQL multimedia and application packages – Part 3: Spatial (2006). https://www.iso.org/standard/38651.html
84. Baumann, P.: Rasdaman – Raster Data Manager, https://zenodo.org/record/1040170, (2017). https://doi.org/10.5281/ZENODO.1040170
85. Obe, R.O., Hsu, L.: PostGIS in Action. Manning, Greenwich (2011)
86. Shekhar, S., Chawla, S.: Spatial Databases: A Tour. Prentice Hall, Upper Saddle River (2005)
87. Rigaux, P., Scholl, M., Voisard, A.: Spatial Databases: With Application to GIS. Morgan Kaufmann, San Francisco (2002)
88. Nebert, D.D.: The SDI cookbook (2004). http://gsdiassociation.org/images/publications/cookbooks/SDI_Cookbook_GSDI_2004_ver2.pdf, Accessed 2 May 2022
89. rasdaman: Home page (2019). http://www.rasdaman.org, Accessed 2 May 2022
90. Wikipedia contributors: Array DBMS (2019). http://en.wikipedia.org/wiki/Array_DBMS, Accessed 2 May 2022
91. Wikipedia contributors: Data cube (2019). http://en.wikipedia.org/wiki/Data_cube, Accessed 2 May 2022
92. Baumann, P.: Language support for raster image manipulation in databases. In: Göbel, M., Teixeira, J.C. (eds.) Graphics Modeling and Visualization in Science and Technology Beiträge zur Graphischen Datenverarbeitung. pp. 236–245. Springer, Berlin, Heidelberg (1993)
93. Baumann, P., Dehmel, A., Furtado, P., Ritsch, R., Widmann, N.: Spatio-temporal retrieval with RasDaMan. In: VLDB'99, Proceedings of 25th International Conference on Very Large Data Bases Edinburgh, 7–10 September 1999. Morgan Kaufmann, San Francisco (1999)
94. Baumann, P., Feyzabadi, S., Jucovschi, C.: Putting pixels in place: a storage layout language for scientific data. In: Proc. IEEE ICDM Workshop on Spatial and Spatiotemporal Data Mining (SSTDM'10) Sydney, 14 Dec 2010. (2010)
95. Aiordachioaie, A., Baumann, P.: PetaScope: An Open-Source implementation of the OGC WCS Geo Service standards suite. In: Proceedings of the 22nd International Conference on Scientific and Statistical Database Management (SSDBM '10) Heidelberg. (2010)
96. International Organization for Standardization: ISO IS 9075-15: Information Technology – Database Languages – SQL – Part 15: Multi-Dimensional Arrays (SQL/MDA) (2018)
97. Misev, D., Baumann, P.: Extending the SQL array concept to support scientific analytics. In: Proceedings of the 26th International Conference on Scientific and Statistical Database Management (SSDBM '14) Aalborg, 30 June–2 July 2014. (2014). No. 10
98. Baumann, P.: The OGC web coverage processing service (WCPS) standard. Geoinformatica **14**(4), 447–479 (2010)
99. Baumann, P., Misev, D., Merticariu, V., Pham Huu, B.: Datacubes: towards space/time analysis-ready data. In: Doellner, J., Jobst, M., Schmitz, P. (eds.) Service Oriented Mapping – Changing Paradigm in Map Production and Geoinformation Management Springer Lecture Notes in Geoinformation and Cartography. (2018)
100. Misev, D., Yu, J., Baumann, P., Campalani, P.: Finding my CRS: A systematic way of identifying CRss. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12) Redondo Beach, 6–9 November 2012. (2012)
101. GeoBlacklight contributors: GeoBlacklight (version 3.4). Zenodo (2022). https://doi.org/10.5281/ZENODO.5851664
102. Durante, K., Hardy, D.: A metadata schema for geospatial resource discovery use cases. Code4Lib J. **25** (2014)
103. OpenGeoMetadata contributors: OpenGeoMetadata (2019). https://github.com/OpenGeoMetadata, Accessed 2 May 2022
104. Battistta, A., Balough, S., Hardy, D., Majewicz, K.: Consortial geospatial data collection: toward standards and processes for shared GeoBlacklight metadata. J. Libr. Metadata **17**(3-4), 183–200 (2017)
105. Wessel, P., Luis, J.F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W.H.F., Tian, D.: The Generic Mapping Tools version 6. Geochem. Geophys. Geosyst. **20**, 5556–5564 (2019). https://doi.org/10.1029/2019GC008515
106. Wessel, P., Luis, J.F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W.H.F., Tian, D., Jones, M., Esteban, F.: The Generic Mapping Tools version 6.3.0. Zenodo (2021). https://doi.org/10.5281/ZENODO.3407865
107. Wessel, P., Smith, W.H.F.: Free software helps map and display data. Eos Trans. AGU **72**(41), 441 (1991)
108. Wessel, P., Smith, W.H.F.: New version of the Generic Mapping Tools released. Eos Trans. AGU **76**, 329 (1995)
109. Wessel, P., Smith, W.H.F.: New, improved version of Generic Mapping Tools released. Eos Trans. AGU **79**(47), 579 (1998)
110. Tozer, B., Sandwell, D. T., Smith, W. H. F, Olsen, C., Beale, J. R., Wessel, P.: Global bathymetry and topography at 15 arc

seconds: SRTM15+V2.0., Earth and Space Sciences 6(10): 1847–1864 (2019)
111. Wessel, P., Smith, W.H.F.: A global, self-consistent, hierarchical, high-resolution shoreline database. J. Geophys. Res. **101**(B4), 8741–8743 (1996)
112. Wessel, P., Smith, W.H.F., Scharroo, R., Luis, J.F., Wobbe, F.: Generic Mapping Tools: Improved version released. Eos Trans. AGU **94**(45), 409–410 (2013)
113. Wessel, P., Matthews, K.J., Müller, R.D., Mazzoni, A., Whittaker, J.M., Myhill, R., Chandler, M.T.: Semiautomatic fracture zone tracking. Geochem. Geophys. Geosyst. **16**(7), 2462–2472 (2015)
114. Wessel, P., Luis, J. F., Uieda, L.: The Generic Mapping Tools version 6.3.0, (2021). https://doi.org/10.5281/zenodo.5708768 (last visited 6.1.2022)
115. Wessel, P., Luis, J.F.: The GMT/MATLAB toolbox. Geochem. Geophys. Geosyst. **18**, 811–823 (2017)
116. Uieda, L., Tian, D., Leong, W. J., Schlitzer, W., Toney, L. Grund, M., Jones, M., Yao, J., Materna, K., Newton, T., Anant, A., Ziebarth, M., Wessel, P.: PyGMT: A Python interface for the Generic Mapping Tools (version v0.4.0). Zenodo. https://doi.org/10.5281/zenodo.4978645 (2021)
117. Luis, J.F., Wessel, P.: A Julia wrapper for the Generic Mapping Tools. In: AGU Fall Meeting (2018). Abstract NS53A-0563
118. Amante, C., Eakins, B.W.: ETOPO1 1 arc-minute global relief model: Procedures, data sources and analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, Boulder (2008)
119. Müller, R.D., et al.: Ocean basin evolution and global-scale plate reorganization events since Pangea breakup. Ann. Rev. Earth Planet. Sci. **44**(1), 107–138 (2016)
120. Antonello, A., Franceschi, S.: Geopaparazzi. Zenodo (2022). https://doi.org/10.5281/ZENODO.5881915
121. Stankovic, J., Neteler, M., Flor, R.: Mobile wireless GRASS GIS for handheld computers running GNU/Linux. Trans. GIS **8**(2), 225–233 (2004)
122. De Donatis, M., Antonello, A., Lanteri, L., Susini, S., Foi, M.: BeeGIS: A new open source tool for mobile GIS applications. Rend. Online Soc. Geol. Italiana **3**, 1–3 (2008)
123. https://www.linux-netbook.com/smartphone-os-market-share/, data by Gartner Inc., Accessed 2 May 2022
124. SpatiaLite: Home page (2020). https://www.gaia-gis.it/fossil/libspatialite/index
125. mapsforge: Presentation at barcamp droidcon 2010 (2010). https://github.com/mapsforge/mapsforge/blob/master/docs/presentations/barcamp%20droidcon%202010.odp, Accessed 2 May 2022
126. GitHub: MBTiles specification (2018). https://github.com/mapbox/mbtiles-spec, Accessed 2 May 2022
127. Antonello, A., Franceschi, S.: GSS Geopaparazzi Survey Server. Zenodo (2022). https://doi.org/10.5281/ZENODO.5882266
128. gvSIG Association: gvSIG online (2009). http://www.gvsig.com/en/products/gvsig-online, Accessed 2 May 2022
129. Geoanalytic: A reference server for Geopaparazzi cloud profiles (2019). https://geoanalytic.github.io/a-reference-server-for-geopaparazzi-cloud-profiles/, Accessed 2 May 2022
130. Antonello, A., Franceschi, S.: SMASH. Zenodo (2022). https://doi.org/10.5281/ZENODO.5882070
131. Christen, M.: Geopython conference (2019). http://www.geopython.net, Accessed 2 May 2022
132. Python Software Foundation: Home page (209). https://www.python.org, Accessed 2 May 2022
133. Erickson, J., Cort, D., Payne, M.: Python GDAL/OGR cookbook 1.0 documentation (2013). https://pcjericks.github.io/py-gdalogr-cookbook, Accessed 2 May 2022
134. PROJ: Home page (2019). https://proj.org, Accessed 2 May 2022
135. QGIS Development Team: QGIS (2019). https://www.qgis.org, Accessed 2 May 2022
136. Story, R.: Folsium 0.8.3 documentation (2013). http://python-visualization.github.io/folium, Accessed 2 May 2022
137. Mapbox: Rasterio: Access to geospatial raster data (2018). https://rasterio.readthedocs.io/en/latest, Accessed 2 May 2022
138. Gillies, S., Wurster, K., Butler, H., Tonnhofer, O., Arnott, J.: Toblerity project: Fiona (2019). https://github.com/Toblerity/Fiona, Accessed 2 May 2022
139. Python Software Foundation: Python interface to PROJ (2006). https://pypi.org/project/pyproj/, Accessed 2 May 2022
140. GeoPandas Developers: GeoPandas 0.10.2 (2022). http://geopandas.org, Accessed 6 Jan 2022
141. Folium Development Team: Python data. Leaflet.js maps (2019). https://github.com/python-visualization/folium, Accessed 2 May 2022
142. Ingargiola, A., et al.: Jupyter/IPython notebook quick start guide (2015). https://jupyter-notebook-beginner-guide.readthedocs.io, Accessed 2 May 2022
143. German Aerospace Center: CODE-DE – Copernicus Data and Exploitation Platform – DE (2022). https://code-de.org/en, Accessed 21 Feb 2022
144. Planet Labs: Interactive notebooks from Planet Engineering (2017). https://github.com/planetlabs/notebooks, Accessed 2 May 2022
145. McKenna, J., MapServer PSC: MapServer. Zenodo (2021). https://doi.org/10.5281/ZENODO.5842012
146. MapServer Development Team: MapServer Fast and customizable Web mapping engine, written in C (2019). https://github.com/MapServer, Accessed 2 May 2022
147. Open Source Geospatial Foundation: GeoServer (2014). http://GeoServer.org, Accessed 2 May 2022
148. Aime, A., Barsballe, T., Garnett, J., NielsCharlier, Romagnoli, D., Smith, K., Fabiani, A., Caradoc-Davies, B., Mbarto, Oliveira, N., Hards, B., Taba90, Pumphrey, M., Roldan, G., Turton, I., Prins, M., Gasdorf, F., Ikeoka, S., Miño, F., Deoliveira, J., Tucker, D., Angreani, R., Otero, J.M., , Imran, Savazzi, D., Tajariol, E., N-Lagomarsini, Hughes, J., Smythe, P., Harrtell, B.: geoserver/geoserver: Zenodo (2022). https://doi.org/10.5281/ZENODO.5854561
149. Open Geospatial Consortium: Web map service (2019). http://www.opengeospatial.org/standards/wms, Accessed 2 May 2022
150. Open Geospatial Consortium: Web feature service (2019). http://www.opengeospatial.org/standards/wfs, Accessed 2 May 2022
151. Open Geospatial Consortium: Web coverage service (2019). http://www.opengeospatial.org/standards/wcs, Accessed 2 May 2022
152. GeoNode contributors: GeoNode (2012). http://geonode.org, Accessed 2 May 2022
153. Open Source Geospatial Foundation: API details – GeoServer 2.14.x user manual (2018). http://docs.GeoServer.org/stable/en/user/rest/api/details.html, Accessed 2 May 2022
154. Ubuntu manuals: debtree – Show relationships between packages (2019). http://manpages.ubuntu.com/manpages/xenial/man1/debtree.1.html, Accessed 2 May 2022
155. PROJ Development Team: News 8.2.1 release notes (2022). https://proj.org/news.html, Accessed 6 Jan 2022
156. Evenden, G.I., Rouault, E., Warmerdam, F., Evers, K., Knudsen, T., Butler, H.: PROJ. Zenodo (2022). https://doi.org/10.5281/ZENODO.5884394
157. Robb, J.M.: A collection of MAPGEN command files suitable for hacking (1989). USGS Open-File Report 89-653. Internet Archive. https://web.archive.org/web/20080528162153/https://pubs.usgs.gov/of/1989/of89-653/mapgen.txt/, Accessed 2 May 2022

158. Wikipedia contributors: Ratfor (Rational Fortran) (2019). https://en.wikipedia.org/wiki/Ratfor, Accessed 2 May 2022
159. International Organization for Standardization: ISO-19111:2019: Geographic Information – Referencing by Coordinates (2019)
160. International Organization for Standardization: ISO 19162:2019: Geographic Information – Well-Known Text Representation of Coordinate Reference Systems (2019)
161. Open Source Geospatial Foundation: GDAL/OGR (2018). https://www.osgeo.org/projects/gdal, Accessed 2 May 2022
162. Rouault, E., Warmerdam, F., Schwehr, K., Kiselev, A., Butler, H., Łoskot, M.: GDAL. Zenodo (2022). https://doi.org/10.5281/ZENODO.5884351
163. GDAL/OGR distributions (2022). https://www.gdal.org/download.html#binaries, Accessed 6 Jan 2022
164. GDAL/OGR contributors: GDAL GeoTIFF file format (2018). https://www.gdal.org/frmt_gtiff.html, Accessed 2 May 2022
165. GDAL/OGR contributors: GDAL data model (2018). https://www.gdal.org, Accessed 2 May 2022
166. Davis, M.: History of JTS and GEOS (2007). http://lin-ear-th-inking.blogspot.com/2007/06/history-of-jts-and-geos.html, Accessed 2 May 2022
167. gvSIG Association: Home page (2009). http://www.gvsig.com/, Accessed 2 May 2022
168. Refractions Research: User-friendly desktop Internet GIS (2005). http://udig.refractions.net/, Accessed 2 May 2022
169. Steiniger, S., Blake, L., Michaud, M.: Open Jump (2018). http://www.openjump.org/, Accessed 2 May 2022
170. GeoServer: Home page (2019). http://GeoServer.org/, Accessed 2 May 2022
171. Open Source Geospatial Foundation: GeoTools the Open Source Java GIS toolkit (2019). http://geotools.org/, Accessed 2 May 2022
172. Centre national de las recherché scientifique: A spatial extension for the H2 database (2013). http://www.h2gis.org/, Accessed 2 May 2022
173. HydrologGIS: Geopaparazzi reference material (2017). https://www.geopaparazzi.org/geopaparazzi/, Accessed 2 May 2022
174. PostGIS Developer Team: PostGIS – Spatial and geographic objects for PostgreSQL (2001). https://postgis.net/, Accessed 2 May 2022
175. GDAL/OGR contributors: GDAL – Geospatial data abstraction library (2018). http://gdal.org/, Accessed 2 May 2022
176. Eclipse Foundation: LocationTech (2019). https://www.locationtech.org/, Accessed 2 May 2022
177. Wikipedia contributors: Well known text (2019). https://en.wikipedia.org/wiki/Well-known_text, Accessed 2 May 2022
178. Wikipedia contributors: Well known binary (2019). https://en.wikipedia.org/wiki/Well-known_text#Well-known_binary, Accessed 2 May 2022
179. Wikipedia contributors: Geography markup language (2019). https://en.wikipedia.org/wiki/Geography_Markup_Language, Accessed 2 May 2022
180. Aime, A., Brown, A., Fabiani, A., Wöstmann, A., Caradoc-Davies, B., Hards, B., Walker, B., Romagnoli, D., Tucker, D., Winslow, D., Tajariol, E., Miño, F., Gasdorf, F., Roldan, G., Mayo, I., Turton, I., Schneider, I., Deoliveira, J., Erickson, J., Eichar, J., Hughes, J., Garnett, J., Jrbeckwith, Jwood106, Morandini, L., Leslie, M., Mbarto, Mcrmcr, Antonello, A., Pazos, M., Prins, M., N-Lagomarsini, NielsCharlier, Oliveira, N., Parma, A., Costa, S., Angreani, R., Giannecchini, S., Smith, K., Uhrig, S., Savazzi, D., Kalén, M., Taba90, Barsballe, T., Joe, Vick, D., Victor: GeoTools. Zenodo (2022). https://doi.org/10.5281/zenodo.5854653
181. Schulz, R.: Geotools: History (1996). http://old.geotools.org/History_8301.html, Accessed 2 May 2022
182. SeaGIS: The SeaGIS project (history) (2001). http://seagis.sourceforge.net/, Accessed 2 May 2022
183. Antonello, A., Franceschi, S., Riccardo, R.: Hortonmachine. Zenodo (2022).
184. GEOMAJAS Development Team: GEOMAJAS (2019). http://www.geomajas.org/, Accessed 2 May 2022
185. Prunayre, F., Coudert, M.: Spatial extension for Talend (2007). http://talend-spatial.github.io/, Accessed 2 May 2022
186. Romagnoli, D., Giannecchini, S.: Additional plugins and extension for the standard Java ImageIO library (2019). https://github.com/geosolutions-it/imageio-ext, Accessed 2 May 2022
187. Brewer, C.A.: COLORBREWER 2.0 color advice for cartography (2001). http://www.ColorBrewer.org, Accessed 2 May 2022
188. Open Source Geospatial Foundation: GeoTools rich client platform (2019). http://docs.geotools.org/latest/userguide/unsupported/swt/rcp.html, Accessed 2 May 2022
189. Agafonkin, V.: I'm an Open Source enthusiast at Mapbox, the creator of Leaflet and 40+ other JS libraries, and a rock musician. AMA! (2018). https://dev.to/mapbox/im-an-open-source-enthusiast-at-mapbox-the-creator-of-leaflet-and-40-other-js-libraries-and-a-rock-musician-ama-1c7o, Accessed 2 May 2022
190. Agafonkin, V.: How simplicity will save GIS (2018). https://speakerdeck.com/mourner/how-simplicity-will-save-gis?slide=42, Accessed 2 May 2022
191. GitHub: CloudMade – Leaflet (2013). https://github.com/CloudMade/Leaflet, Accessed 2 May 2022
192. Agafonkin, V.: Tutorials – Leaflet – a JavaScript library for interactive maps (2018). https://leafletjs.com/examples.html, Accessed 2 May 2022
193. Open Source Geospatial Foundation: Tile map service specification – OSGeo (2018). http://wiki.osgeo.org/wiki/Tile_Map_Service_Specification, Accessed 2 May 2022
194. Open Geospatial Consortium: Web map service (2018). http://www.opengeospatial.org/standards/wms, Accessed 2 May 2022
195. Gillies, S., Butler, H., Daly, M., Doyle, A., Schaub, T.: The GeoJSON format (2018). https://tools.ietf.org/html/rfc7946, Accessed 2 May 2022
196. Wikipedia contributors: Object-oriented programming (2018). https://en.wikipedia.org/w/index.php?title=Object-oriented_programming&oldid=862368729, Accessed 2 May 2022
197. Snydman, S., Sanderson, R., Cramer, T.: The International Image Interoperability Framework (IIIF): A community & technology approach for web-based images. In: Archiving 2015 Los Angeles, 19–22 May 2015. (2015). Internet Archive. https://web.archive.org/web/20220120153949/, Accessed 2 May 2022
198. Murris, P.: Interview with Patrick Hogan WWJ project manager (2009). http://patmurris.blogspot.com/2009/01/interview-with-patrick-hogan-wwj.html, Accessed 2 May 2022
199. NASA: NASA World Wind (2019). https://worldwind.arc.nasa.gov/, Accessed 2 May 2022
200. Wikipedia contributors: Shuttle radar topography mission (2019). http://en.wikipedia.org/wiki/Shuttle_Radar_Topography_Mission, Accessed 2 May 2022
201. NASA: NASA World Wind Open Source agreement version 1.3 (2019). https://worldwind.arc.nasa.gov/LICENSE.html, Accessed 2 May 2022
202. NASA: NASA World Wind project (2022). https://worldwind.arc.nasa.gov, Accessed 6 Jan 2022
203. GitHub: WorldWind Java – Community Edition (WWJ-CE) (2021). https://github.com/WorldWindEarth/WorldWindJava, Accessed 2 May 2022
204. GitHub: WorldWindJS (2021). https://github.com/WorldWindEarth/worldwindjs, Accessed 2 May 2022

205. GitHub: WorldWindAndroid (2021). https://github.com/WorldWindEarth/WorldWindAndroid, Accessed 2 May 2022
206. https://worldwind.arc.nasa.gov/news/2019-08-21-project-update/, Accessed 18 8 2021
207. Müller, R.D., Cannon, J., Qin, X., Watson, R.J., Gurnis, M., Williams, S., Pfaffelmoser, T., Seton, M., Russell, S.H., Zahirovic, S.: Gplates: Building a virtual earth through deep time. Geochem. Geophys. Geosyst. **19**, 2243–2261 (2018)
208. Cannon, J., Chin, M.: GPlates 2.3.0. Zenodo (2021). https://doi.org/10.5281/ZENODO.5885988
209. Qin, X., Müller, R., Cannon, J., Landgrebe, T., Heine, C., Watson, R., Turner, M.: The Gplates geological information model and markup language. Geosci. Instrum. Methods Data Syst. **2**, 365–428 (2012)
210. International Organization for Standardization: ISO 19136-1: Geographic information – Geography Markup Language (GML) – Part 1: Fundamentals (2020)
211. Wessel, P., Smith, W.H.F.: New, improved version of generic mapping tools released. Eos Trans. AGU **79**(47), 579–579 (1998)
212. Cannon, J., Lau, E., Müller, R.: Plate tectonic raster reconstruction in Gplates. Solid Earth **5**, 741 (2014)
213. Gurnis, M., Turner, M., Zahirovic, S., DiCaprio, L., Spasojević, S., Müller, R.D., Boyden, J., Seton, M., Manea, V.C., Bower, D.: Plate tectonic reconstructions with continuously closing plates. Comput. Geosci. **38**, 35–42 (2012)
214. Gurnis, M., Yang, T., Cannon, J., Turner, M., Williams, S., Flament, N., Muller, R.: Global tectonic reconstructions with continuously deforming and evolving rigid plates. Comput. Geosci. **116**, 32–41 (2018)
215. Emde, A.: OSGeoLive. Zenodo (2022). https://doi.org/10.5281/ZENODO.5884641
216. Bowman, H., Tsotos, A., Mandel, A.: OSGeoLive bootable Linux distro for geospatial application (2019). https://sourceforge.net/projects/osgeo-live/files/13.0/, Accessed 2 May 2022
217. OSGeoLive contributors: OSGeoLive project metrics (2019). https://live.osgeo.org/en/metrics.html, Accessed 2 May 2022
218. Emde, A.: OSGeoLive Documentation. Zenodo (2022). https://doi.org/10.5281/ZENODO.5884859
219. OSGeoLive Project: Documentation archive (2022). https://live-archive.osgeo.org, Accessed 6 Jan 2022

Peter Löwe Peter Löwe is the Chapter Editor. He received his PhD in Geography from the University of Würzburg, Germany (2003). He currently works at the WZB Berlin Social Science Center, focusing on IT Services and e-Science. He joined the German OSGeo local chapter in 2001 and is an elected OSGeo charter member and OSGeo Vice President for Open Geoscience.

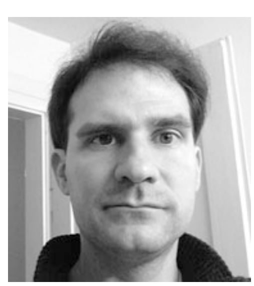
Álvaro Anguix Alfaro Álvaro Anguix Alfaro is currently General Manager of the gvSIG Association. An engineer in topography by education (Polytechnic University of Valencia, 2000), he has developed his professional work around the gvSIG project and free geomatics. He has participated in geomatics projects developed in more than 30 countries.

Andrea Antonello Andrea Antonello, an expert on GeoPaparazzi, JTS, and Nasa WorldWind, received his PhD from the University of Urbino. He is cofounder of HydroloGIS S.r.l., a company that focuses on GFOSS solutions for risk modeling, water management, and digital field mapping. He teaches advanced geomatics at the Free University of Bolzano, and is an OSGeo charter member and a passionate advocate for open science.

Peter Baumann Peter Baumann is Professor of Computer Science, researching on datacube services and their application in science and engineering. With the rasdaman engine, he and his team have pioneered datacubes and Array Databases, with 160+ scientific publications and international patents. The rasdaman datacube engine has received a series of innovation awards. Peter Baumann is Editor of the datacube standards in ISO and OGC, and strongly contributes to INSPIRE.

Mario Carrera Mario Carrera is the Language and Communication Manager at the gvSIG Project and Training Manager at the gvSIG Association. He has a degree in Topography and a Bachelor's degree in Geodesy and Cartography from the Polytechnic University of Valencia. He has been participating in the coordination and implementation of spatial data infrastructures (SDI) in several companies and public administrations around the world.

Kim Durante Kim Durante received her MLIS from the University of Wisconsin-Milwaukee and an MA in Art History from Georgia State University. She works at Stanford University Libraries as the Science Metadata Librarian. Her work includes the curation and management of digital geospatial data. She has written this chapter's section on GeoBlacklight and gdal/ogr.



Marco Hugentobler Marco Hugentobler is an expert on QGIS. He received his PhD from the University of Zurich, Switzerland (2004). He worked for Autoform GmbH and the Institute of Cartography, ETH Zurich. He is now a co-owner of Sourcepole AG, a software company providing consulting and development services for open source GIS.

Steve Lime Steve Lime is Data and Applications Manager at the Minnesota Department of Natural Resources. He is the primary technical developer of the MapServer-based DNR Data Deli, providing significant data to the GIS community. He serves on the Master of Geographic Information Science graduate faculty and has advised a number of students in the program. He also taught a popular course on GIS and the Internet at the University of Minnesota.

Helena Mitasova Helena Mitasova is an expert on GRASS GIS. She received her PhD from the Slovak Technical University in Bratislava (1987). She joined the GRASS GIS development team at the US Army Corps of Engineers Laboratory in Champaign and the University of Illinois in 1991. She is now Professor and Associate Director of Geovisualization at the Center for Geospatial Analytics, North Carolina State University in Raleigh.

Dietmar Müller Dietmar Müller received his undergraduate degree from the University of Kiel, Germany, and his PhD in Earth Science from the Scripps Institution of Oceanography, La Jolla/California in 1993. He is Professor of Geophysics at the School of Geosciences, University of Sydney. With his EarthByte research group he combines tectonics, geodynamics, and sedimentology to build virtual Earth models from global to basin scales. He is an expert on GPlates.

Markus Neteler Markus Neteler, PhD, has been Project Coordinator of GRASS GIS since 1997 and a founding member of OSGeo, GFOSS, and FOSSGIS, as well as cofounder of mundialis, a geospatial analysis and EO business. At mundialis, he oversees remote sensing, GIS, and cloud computing-related projects.

Jack Reed Jack Reed is a software engineer working on digital discovery of and access to cultural heritage content. A cofounder of the GeoBlacklight project and an active contributor to many open source software projects,. He also serves on the Executive Committee of The International Association for Geoscience Diversity. To this chapter he contributed to the sections on GeoBlacklight, GeoServer, and Leaflet.

Christian Strobl Christian Strobl received his PhD in Geology from Munich Ludwig-Maximilians University, where he also teaches GIS. From 2006 on Christian worked as GIS and Remote Sensing Scientist at the German Aerospace Center. Now he is responsible for the coordination of INSPIRE and Copernicus affairs at the Bavarian State Ministry of the Environment. He contributed the sections on PostGIS and GeoPython to this chapter.

Paul Wessel Paul Wessel received his PhD in Geological Sciences from Columbia University, New York, in 1990, and he has been with the Department of Earth Sciences in the School of Earth and Ocean and Technology at the University of Hawaii at Mānoa since 1991. He specializes in lithosphere geodynamics, plate tectonics, and scientific software design. He is an expert on GMT and authored this section in the chapter.

Smart Cities

31

Alan Leidner and George Percivall

Contents

A. Leidner (✉)
President of the NYC Geospatial Information Systems and Mapping Organization or GISMO
New York, NY, USA
e-mail: leidnera@nyc.rr.com

G. Percivall
GeoRoundtable
Crofton, MD, USA
e-mail: percivall@ieee.org

Abstract

Smart cities, the current movement to advance beneficial programs and applications for urban areas, depend heavily on innovative information technologies, the expansion of data types, and increases in data volumes. This chapter explains that spatially enabled systems are critical to the success of smart cities because their unique integrating, analyzing and visualizing capabilities can multiply the power of conventional IT systems. The chapter identifies the characteristics that distinguish spatial systems and discusses how increasing amounts of interoperable, spatially enabled data is becoming available due to new sensor technologies, the Internet of Things (IoT) and social media. The chapter identifies the standards, architecture and organization structure that can create an effective spatial information environment. It points towards the use of spatial technologies in advanced artificial intelligence applications. There is an examination of the measurable benefits made possible by spatial systems including enhanced revenues, lifesaving methods, and increased productivity. The chapter concludes by re-emphasizing the importance of spatial systems for smart cities initiatives.

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6_31

Keywords

CityGML · interoperability · location intelligence · smart city

In this chapter, we will explore the movement referred to as *Smart Cities* from the standpoint of the critical role that spatial systems and capabilities play in this, the latest cycle of the information technology and scientific revolution. According to BSI PAS 180 – Smart Cities Vocabulary [1], A smart city provides effective integration of physical, digital, and human systems in the built environment to deliver a sustainable [2], prosperous and inclusive future for its citizens.

Another definition of smart city puts even more emphasis on information and information technology. "A smart city is an urban area that uses different types of electronic data collection sensors to supply information which is used to manage assets and resources efficiently. This includes data collected from citizens, devices, and assets ..." [3]. Similarly, according to "Smart Cities, Smart Future" by Mike Barlow and Cornelia Levy-Bencheton, "Smart cities are complex blends of interoperable technologies, systems and services designed and orchestrated to help people lead productive, fulfilling, safe and happy lives" [4].

31.1 Introduction

31.1.1 Prelude to Smart Cities

Smart cities are the latest incarnation of past movements that sought to promote societal improvement and used terms like excellence, reengineering, change, quality, innovation, e-government, and digital revolution. Smart cities draw upon these past concepts but reflect our present time when increases in the volume, types and quality of data, and rapid advances in networking and information technologies represents a qualitative and quantitative advance from the past.

Smartness depends upon the human capacity to observe the world, gather information, and process that information into knowledge, insights, decisions, and actions. Naturally, an essential part of comprehending the world is understanding location: where you are, what is around you, where you want to go, and where things important to you are located. As civilization advanced, location became ever more important, guiding explorers, supporting commerce, and organizing living space. Now, with more than half the world's population living in urban areas [5] it is important that we fully appreciate that everything contained and occurring within a city has a location: building footprints, property boundaries, utility networks, the movement of people and vehicles at street level, and all that goes on inside buildings. Only by taking the location of everything into account—in as smart a manner as possible—can we hope to grasp the complexity of urban life and meet its challenges.

We know that smartness did not begin in just the past 5 or 10 years. Looking into the past we see successive waves of technology and innovation that have made life better. But since the end of World War Two, one of the primary means for achieving smartness has been through innovations of computer technology based on the development of digital information, integrated circuits, computing devices, telecommunications, and computing software. Computers, with their ability to create, manage, analyze, and transmit huge volumes of information, are now behind most of the scientific advances and operational improvements that we have witnessed in our lifetimes.

Computer technology revolutionized the way information was recorded, stored, and analyzed. However, the location field of a data record was often the most problematic. There were many different ways of assigning and recording addresses with inconsistency and mistakes more common than not. Address data is known to have an error rate that can exceed ten percent. And addresses could not adequately deal with vacant land, open spaces, water bodies, the underground, or space above buildings. Conventional databases also had no way of either recording or analyzing the geographic relationship of physical objects in space, such as a building's placement on its parcel and the access points of utilities that supplied it with water, electricity, and other services. While the large majority of data records had a location field, bringing information together on the basis of location was unreliable, when it was not impossible.

It took geographers and data scientists many years before the indicators of location could be digitally defined and automated to form the basis for spatial information systems. The moment of birth for geographic information systems (GIS) is tied back to the work of Dr. Roger Tomlinson, who developed the first known computerized spatial database for the Canada Land Inventory in the early 1960s. A revolutionary advance occurred when a network of global position satellites was put into orbit, in the 1970s, which enabled Earth coordinates to be rapidly and accurately captured for any point on Earth [6]. During this time of spatial evolution, practitioners understood that spatial systems were evolving as an extension of information technology; but that adding the location element created a new dimension of smartness, often referred to as location intelligence. Early applications of GIS were standalone systems, residing in agency silos, that served limited functions within environmental, planning, health, transportation, and public safety agencies. Often, individual agencies developed a basemap customized for their unique needs. Data represented on one map often could not be used with the data from another. Spatial systems were created, but they were not interoperable.

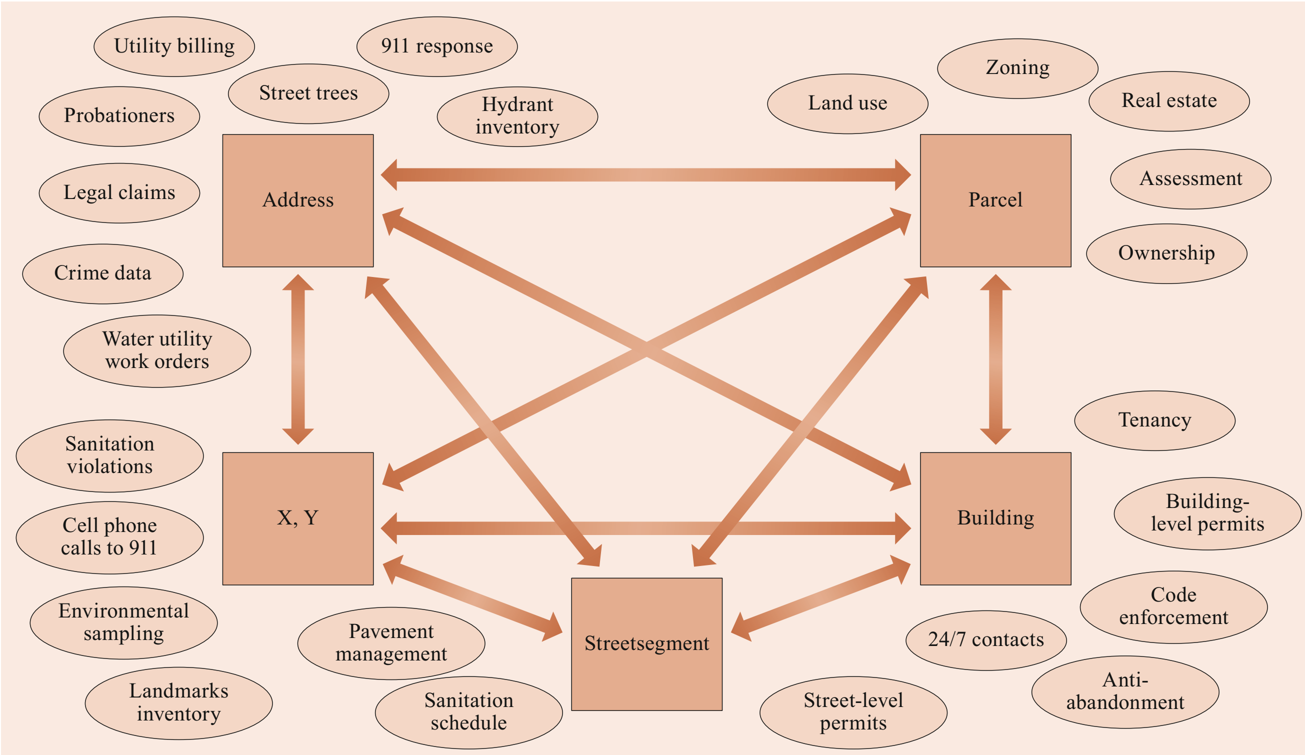


Fig. 31.1 Geospatially enabled enterprise data integration (GEDI). Aligned geographic signifiers allow data integration and interoperability

31.1.2 Enterprise GIS

From the mid-1980s through the 1990s, many large municipalities and urban counties advanced from standalone spatial systems to Enterprise GIS (Fig. 31.1). To achieve this many jurisdictions, such as New York City, created photogrammetric basemaps, deriving streets, buildings, and other identifiable objects and boundaries from aerial photography. These foundation or framework layers created locational anchors for other datasets, allowing them to be stacked one on top of the other for visualization and analysis. This made it possible for the now standardized location fields of databases to be used as the common element to conduct searches and analyses. Enterprise spatial systems could grow to encompass hundreds of datasets from dozens of participating municipal agencies. Enterprise spatial systems enabled increasing levels of data integration, analysis, and … smartness. Decision support became more robust, and city operations from emergency response to responding to pothole complaints, and hundreds of other applications as well, benefited from more accurate, comprehensive and timely information organized by location.

During the late 1990s enterprise GIS allowed government and private firms using spatial capabilities to get smarter. At the same time, it became increasingly clear that information technology in general was having a revolutionary effect on how cities functioned. In the early 2000s, the idea of Smart Cities started to take hold as a term to characterize the growing benefits being achieved by IT systems in general.

If asked to pick a date when the concept of *spatially enabled smart cities* originated, we would choose September 11, 2001, the day of the attack on the World Trade Center in NYC. The formation of the Emergency Mapping and Data Center (EMDC) by NYC's Office of Emergency Management demonstrated the importance of spatially enabled enterprise data integration to support the response to a massive disaster event. Thousands of maps and spatial analytic products were provided to responders from dozens of organizations, and those organizations contributed their data so it could be pooled together for analytic purposes. Sensing technologies on a large scale were utilized. GPS-enabled data collection tools were deployed to the field. Interactive maps were posted to public websites to inform the public about critical aspects of the response effort. Underground infrastructure data from all utility organizations was assembled, integrated and analyzed (Fig. 31.2) (for more details, see Case Study of 9/11 in Sect. 31.5.5). Although getting all that data to work together was a time-consuming, technical challenge that was realized imperfectly, those involved fully recognized that the power and benefits of spatially enabled data interoperability would be fundamental to any future idea of smartness.

Over the past 20 years we have witnessed a rapid transformation of our technological environment. In addition to the

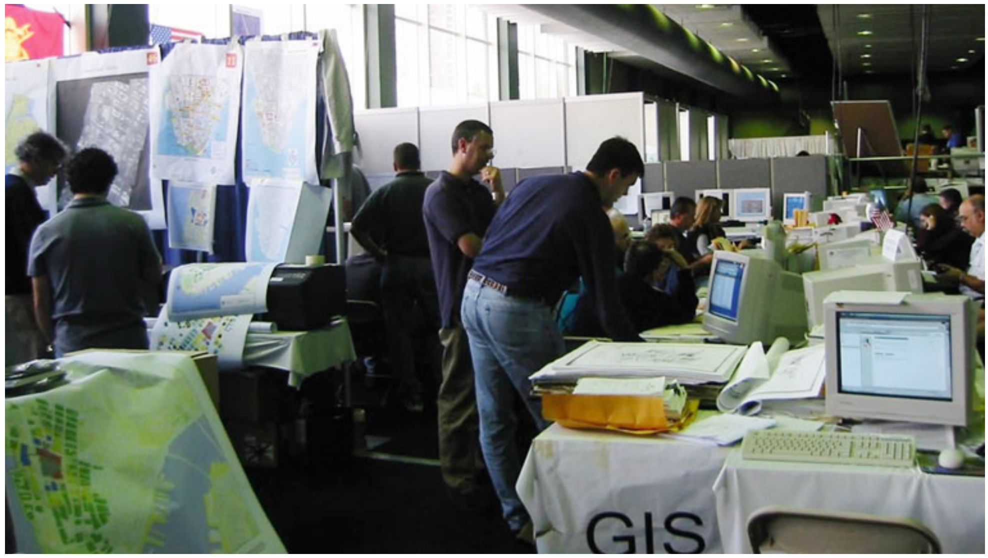


Fig. 31.2 Spatial support provided by the NYC Emergency Mapping and Data Center (EMDC) in response to 9/11. (Courtesy of NYC Department of Emergency Management)

increases in digital storage and computing capacities, many new types of computer devices and sensor technologies have gained widespread use. Almost everyone now carries a smart phone equipped with a GPS receiver, and GPS-enabled wearables are increasing in popularity. At the same time sensors, collecting and transmitting many kinds of spatially enabled data, have increased in functionality and capacity. All these new data sources—from people and things—are tied together by the internet and by a common spatial framework, creating the potential for bringing increasing volumes and types of data together, at enormous speeds, for uses never before contemplated.

31.1.3 Spatial Data Infrastructures

As cities, counties, and states were starting down the path to enterprise GIS, the US Federal government and international standards setting bodies were working to create national spatial frameworks in which local spatial infrastructures could fit. In April 1994, US Executive Order 12906 codified the coordination of Federal, State, local, and tribal governments, and the private sector, to be on a standardized National Spatial Data Infrastructure (NSDI) to support public and private sector applications of geospatial data in such areas as transportation, community development, agriculture, emergency response, environmental management, and information technology.

The goal as described in the Global Spatial Data Infrastructure (GSDI) Cookbook [7] was to be able to access, integrate, and use spatial data from disparate sources in guiding decision-making. The ability to make sound decisions collectively at the local, regional, and global levels is improved by implementation of SDI for compatibility and interoperability across jurisdictions. Through common conventions and technical agreements decision-makers can more readily discover, acquire, exploit, and share geographic information vital to the decision process. The use of common conventions and technical agreements makes economic sense because they limit the cost involved in the integration of information from various sources and eliminate the need for parallel and costly development of tools for discovering, exchanging, and exploiting spatial data.

Development of NSDI by the Federal Geographic Data Committee (FGDC) involved promoting development and eventual endorsement of open standards from the International Organization for Standardization (ISO) Technical Committee 211 (Geographic information/Geomatics) and the Open Geospatial Consortium (OGC). SDI development was a primary reason for the formation of the OGC in 1994. ISO and OGC standards are essential elements in SDIs around the world. This is because SDIs are data and service networks, and networks depend on open standards. Making policies that maximize the use of geospatial products, solutions, and services that implement these standards, is the best way to maximize returns on investment.

While SDI implementation was more pervasive at state and federal levels [8], the role of data sharing interoperability based on standards remains an imperative for local governments. In the early 2000s, a hybrid or "middle-out" approach began to emerge that took combined elements of NSDI with the dynamics of local government [9]. By recognizing the value of SDIs based on open standards to local governments, NYC again led the way. To address the challenges encountered in the 9/11 response, NYC GIS and OGC initiated a discussion of the coordinated development of open standards to meet urban interoperability needs. Several OGC Testbeds were based on needs and scenarios from NYC to drive interoperability solutions. OGC Testbeds are the core of the OGC Innovation Program, resulting in rapid definition

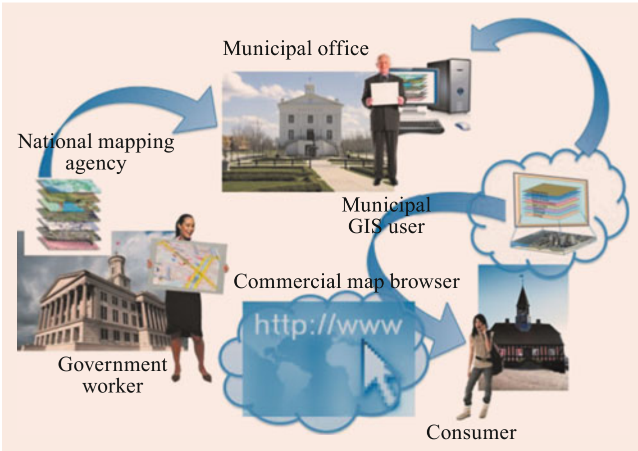


Fig. 31.3 Spatial Data Infrastructure brings distributed data to consumers. (© OGC [10])

and implementation of open consensus standards in OGC's Standards Program (Fig. 31.3).

31.1.4 Key Characteristics of Information-Enabled Future Smart Cities/Societies

The spatial components for Smart Cities build on Enterprise GIS and SDI, but more must be done. Essential will be the continued work of several international standards development organizations: ISO, IEC, ITU, OGC, and others, which broadly address the data and technology context of Smart Cities. The future information environment will be one where everyone and everything is capable of providing and receiving multiple kinds of information, most of it containing a location identifier. The richness of this data makes it possible to imagine ever increasing levels of smartness based on the following characteristics:

- **Data accuracy, completeness and timeliness**: the evolving data and technological environment makes it possible to improve the data available to us for smart purposes. Smart data is data that is accurate and detailed and made available to users when and where they need it.
- **Data interoperability:** the data from traditional IT systems tends to be isolated in silos. It remains a challenge to bring different kinds of datasets together even as it is recognized that data interoperability enables new insights and knowledge. Smart Cities demand that data in any combination needed to solve a problem or support an operation be quickly and easily brought together for analysis and sharing.
- **Improved analytics**: with so many new, high-volume sources of data becoming available, Smart Cities must develop improved analytic techniques and methods to turn combinations of interoperable data into increasingly useful intelligence products. Also, given the large amounts of unstructured data made available via social media, personal devices, and the Internet of Things, new methods need to be found to process this data in ways that allow the extraction of useful information using big data analytics.
- **Suitability for artificial intelligence (AI)**: high volumes of data, such as those generated in response to a large-scale disaster event, can be overwhelming. AI is a branch of computer science characterized by the use of techniques that automatically identify, combine, extract, analyze, and process information, creating intelligent products for a variety of users, from workers in the field to executive suite decision-makers. AI systems—through a process called machine learning—are being designed to use the feedback from errors and mistakes to improve future performance. For example: computer models that predict the path and intensity of hurricanes can incorporate lessons learned from past inaccuracies so that future forecasts are more precise.
- **New, real-time sources:** web technologies encourage the easy publication and immediate access to data coming from many sources. Social media allows individuals to post information that immediately affects events. Twitter is regularly used by protestors and demonstrators in order to communicate and affect events. All kinds of sensors communicate the situation in urban environments to anyone on the web. The Internet of Things already has more connected devices online that there are people in the world.
- **Automation support:** improved sensor, processing, and communications technologies can capture, transmit, and analyze high volumes of spatially enabled data at split second speeds, enabling robots and other kinds of mechanical equipment, which move in three-dimensional space, to become smarter. For example, we are now witnessing the evolution of smart vehicles that are increasingly performing driver-assist functions and may, in the future, take over the entire job of driving. This will require continuous, real-time awareness of vehicle position and the location of nearby people and objects, in order to move safely from one place to another.

The above are key characteristics of data and technology that will support Smart Cities programs in the future. In the next sections, we will show how these capabilities are empowered by spatially enabled applications and systems, utilizing data whose spatial characteristics can serve as the glue linking different kinds of data together. These spatially empowered systems have the ability to multiply the power of conventional IT applications.

31.2 Unique Capabilities of Spatially Enabled Systems

Maps are the most recognizable and readily accessible spatial information. Yet, a focus on maps often obscures the fact that the power of location goes well beyond visualization. To this day, it remains common to hear talk about *big data analytics* without mention of spatial systems that have been integrating and analyzing combinations of large, location enabled datasets for decades. When most people look at a 3-D semantic model of an urban landscape, they do not understand that every pixel (or voxel) and every object represented is both highly accurate and intelligent: the carrier of attribute data that can be analyzed in an almost infinite variety of ways.

So what is it exactly that makes spatially enabled systems so powerful and useful? What follows is a partial listing of spatial capabilities that extend the power of conventional IT systems. Members of the New York State GIS Association at their September, 2018 Summit, when surveyed informally, stated unanimously that the addition of spatial capabilities at least doubles the power of nonspatial information systems. The implications for Smart Cities are enormous: These spatial powers enhance all the key aspects of data usage identified in Sect. 31.1.4, which characterize the use of information technology for Smart Cities programs.

31.2.1 Precise Location – Instruments for Large-Scale Data Gathering

Determining the exact location of points on the ground was once the sole domain of surveyors and navigators employing manual and semimechanical methods to provide geospatial coordinates such as latitude and longitude. In 1967, the US government began launching the Timation satellites that transmitted signals allowing anyone with the proper equipment to calculate their location. This was the precursor to the now ubiquitous GPS network [6]. Location points, assembled as a map-matrix of pixels could then be differentiated to provide spatial identity to lines, polygons, and areas such as streets, buildings, and natural features. With the addition of vertical data, now obtained from lidar sensors, pixels can be turned into voxels, the basic building blocks for representing objects in three dimensions. Other than being the passive carrier of spatial coordinates, conventional IT systems have no ability to use location information for mapping and analytic purposes. Generic databases must be augmented with spatial commands that can measure distance between two points, tell whether two objects touch, and know which way water flows in a network or pipes, and traffic moves on a street grid. It took the development of geospatial information systems to make this possible.

31.2.2 Geocoding Application to Integrate Location Identifiers

Before the advent of GPS, many types of objects and areas, common to cities, were given individual identities that placed them at a particular location and provided the basis for maps created on paper or other recording media. For example, buildings had street addresses and unique building identification numbers. Every property parcel had a block number and lot number. Census tracts were defined by street centerline segments. With the ability to give standardized spatial coordinates to these objects, it became possible to relate them to each other with a high degree of spatial accuracy. Thus, data tied to buildings could be related to data tied to parcels and to surrounding streets. Within a city, or any other kind of jurisdiction, this meant that all data related to a spatial object could be integrated and made interoperable.

31.2.3 Building Multijurisdictional and National Data Layers

Conceptually, building framework data layers, considered to be essential for enterprise GIS, that are seamless across jurisdictional borders and can be assembled into national coverages, is simple. See Fig. 31.4 which shows the interconnection of Lego blocks. This illustrates that once key components of data layers are spatially standardized, they can be layered and connected in any way desired across areas as large as needed. In this way, information from federal, state, local, private sources can be put together as long as each level of government conforms to common standards. GIS framework layers, suitable for standardization include imagery, elevation, hydrography, parcels, streets and addresses, and buildings. Two or more jurisdictions can then link their data across boundaries making sure all features are properly aligned and connected. Of course, achieving this level of this integration is not easy – especially because governments lack the awareness and determination for national standard setting. However, we see hopeful signs.

- The National States Geographic Information Council (NSGIC: www.nsgic.org) made up of the Chief Geospatial Information Officers of all U.S. states has made progress with their "For the Nation" proposals, and is working with the Federal Geospatial Data Committee (FGDC: www.fgdc.gov) composed of Federal agency GIO's, to fund the development of key national framework layers.
- A number of states contract for the collection of statewide imagery, providing it at low cost to their counties and municipalities, which are given the opportunity to upgrade the resolution of the product to suit their needs,

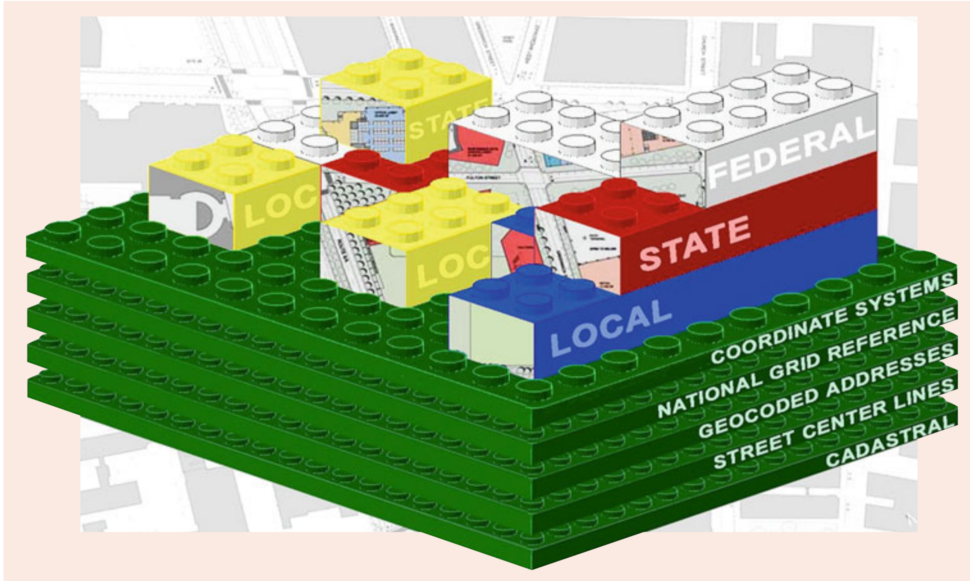


Fig. 31.4 Vision for integrated federal, state, and local data. When data is created by different levels of government on common foundation layers to common standards, it can be combined across geographic areas

within the framework of common accuracy standards. New York State in particular has created statewide layers for imagery, streets, parcels, and addresses.

- Similar to states, a number of counties, such as Westchester, Rockland, and Suffolk Counties in NYS, collect county wide data and make it available to their municipalities.
- Since Westchester and New York City independently created their enterprise GIS data to a common standard, it only requires the determination to properly fit features along their border for their GIS data to be fully aligned.
- The Federal Homeland Infrastructure Foundation Layer Data program of the U.S. Department of Homeland Security has assembled more than 600 national infrastructure datasets (HIFLD Open Data (arcgis.com)) many of which are now available for public download and use.

We hope these steps forward both in the U.S. and other nations continue. The full benefits of GIS for innumerable applications that save lives, improve services, protect the environment, can only be achieved when data across jurisdictions and nations can be made interoperable.

31.2.4 Giving Location Identity to Sensors

Sensors are devices that capture and transmit observations about their surroundings—the characteristics of natural and human phenomena—and are used to detect and measure microorganisms, chemicals, temperature, vibration, speed, and even human presence and behavior. Sensor observations can be used to trigger automated mechanical devices and to send alerts should dangerous conditions be detected. A vital attribute of information collected from a sensor is its location, because measurements are meaningless without knowing where they came from and without being able to synthesize readings from across a larger area to calculate variation and detect patterns. Hence, practically all sensors have the ability to impart location identity to the information they collect and transmit. Additionally, the data from different kinds of sensors covering the same area can be made interoperable if location data standards are adopted. The fusion and analysis of data from multiple types of sensors can give a richer understanding of an area under observation or study.

31.2.5 Spatial Queries and Linked Data

SQL, or structured query language, is the way software programmers ask questions of databases and perform analyses. Spatial SQL expands the range of conventional SQL by enabling queries to address spatial points, lines, and objects and the attribute data associated with them. Spatial SQL, unlike standard SQL, can determine the distances between points and objects, and whether objects are touching, overlapping or in any other kind of spatial relationship with each other. Spatial SQL substantially increases analytic options and enables the achievement of new levels of intelligence to support government and business operations and decision-making. The use categories and use cases described in Sect. 31.5 of this chapter all take advantage of the power of Spatial SQL. One notable application is the use of CompStat by police departments across the country to map crime incidents, identify and analyze crime patterns by incident type, location and time of day; and use this information to design effective strategies (Sect. 31.5.5).

Linked Data [11] is defined as a method of publishing structured data so that it can be interlinked and become more

31

useful through semantic queries [12] across the web. Work is now being done to extend Linked Data to include spatial and temporal characteristics so that, for example, Linked Data in Resource Description Framework (RDF) [13] or Web Ontology Language (OWL) [14] can be queried by using GeoSPARQL [15], a web query language. In this way, structured and semistructured location enabled data can be mixed, exposed, and shared across different applications. Part of the vision of spatially enabled Linked Data is for the internet to become a global database that can serve as a platform to extend spatial analytic capabilities.

31.2.6 Network, Routing, and Analysis

The built environment is composed of physical elements such as pipes, conduits, tunnels, and valves that are networked together to form underground utility systems and building infrastructure. The various products flowing through these systems include water, sewerage, electric power, gas, and data. Spatial systems have the intelligence to know which network components are connected to each other, where they are located, and in which direction network products are traveling. Spatial systems also give locational identity to the sensing and controlling devices that allow such networks to be monitored and actively managed. These capabilities make possible a wide range of asset management applications that can give operators and owners important tools with which to operate and maintain their networks.

Networking and navigation: the networking capabilities of spatial systems can also be applied to transportation routes that can be mapped in detail, whether for travel underground, on roadways, across waterways, and through the air. The economy of our planet depends on the supply chains sending and receiving raw materials and finished products, and spatial systems have the ability to optimize and monitor supply chains and transportation routes. GPS-enabled sensors can track the location and characteristics of all vehicles in real time. The fact that every street and roadway in the US is mapped, with details about traffic direction and capacity, as part of a connected network, enables vehicle fleets to be efficiently routed and has made it possible for emergency response units, and new taxi services, to go from where they are to where they need to go in the most efficient way possible.

31.2.7 Spatial Visualization: Maps and 3-D

Conventional IT systems visually represent data in the form of graphs and charts. Spatial databases additionally allow data to be represented in the form of maps. Spatial systems have the ability to provide operators with the choice of displaying any combination of hundreds of spatial layer options, with the assurance that the different data layers will accurately represent relationships between different kinds of objects. While spatial analytics is invaluable, we should not discount the importance of the ability of the human mind to identify patterns from data represented on a map. Map visualizations and spatial analytics work synergistically: operators notice patterns by visual inspection and then turn to spatial analytics to confirm or expand upon their intuitive understandings. Additionally, a map is a very effective graphical user interface (GUI) that gives technical and nontechnical personnel alike the ability to maneuver through large amounts of data to discover and interact with information. Many government and business online applications utilize maps as a primary means of guiding citizens and customers through a wide variety of transactions.

The world is not a flat map, and map projections that create various kinds of flat two-dimensional maps will always be limited, even though they will be with us for some time to come because they are embedded in our legal definitions of boundaries and locations. We already see how to go beyond flat map projections based on digital data models and computer visualization. Google Earth was successful because computers became capable of 3-D visualization. The next generation of 3-D visualization capabilities, based on open standards, is being implemented with indexed 3-D scene layers (i3S) [16], now released as OGC Community Standards. Data analytics on the globe have become easier with the advent of Discrete Global Grid Systems (DGGSs) [17]. DGGSs do not have the troublesome attributes of date lines and poles that come with map projections. Analytics in DGGS provide capabilities that are revolutionary.

The evolving capabilities described above are unique to spatial systems and provide a significant boost to Smart Cities applications. As increasing amounts and types of spatially enabled data become available, and as new spatial technologies increase capacities and capabilities, IT systems will benefit from extra measures of spatial power to pull all this data together and make sense of it.

31.3 Spatial Data Options for Smart Cities

31.3.1 The Spatially Enabled Data Explosion

Around the year 2000, we entered another phase in both the IT and the spatial revolutions. Since then, we have experienced a quickening of the growth of data volumes spurred by new types of location enabled data, new applications, and technologies that can make everyone a data user and everything a data generator (Fig. 31.5). The availability of these growing data resources is creating enormous opportunities for ever more effective smart applications. These developments are greatly aided by the fact that most of this data is

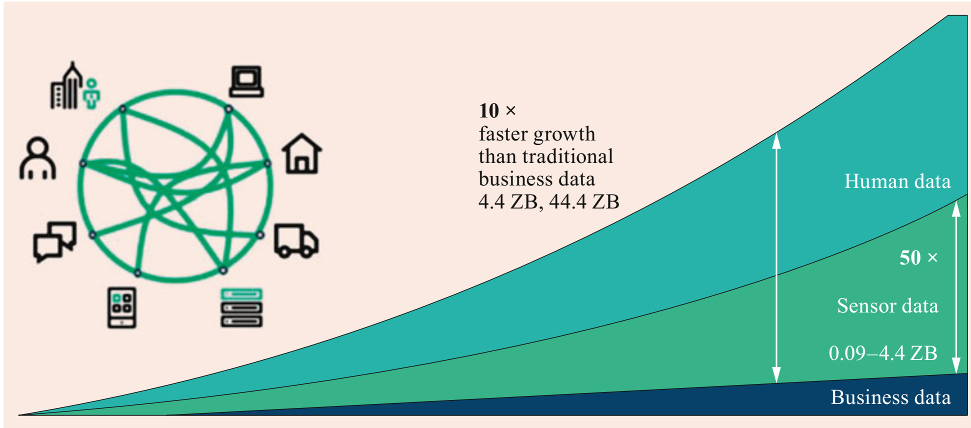


Fig. 31.5 Expected growth rate of data. After [18]. Courtesy of Hewlett Packard Enterprise. A ZB is a zetabyte which is a large measure of digital data volume

being tagged with location information enhancing the potential for interoperability, and for more robust, spatially driven analytics. Below are some of the contributors to this new, richer spatial environment.

- **Smart phones and wearables**: smart phones, smart watches, and other devices carried by individuals have changed the information technology environment as much as desktop computing did in the 1980s. Smart phone ownership now extends to more than two billion people worldwide, and it is easy to foresee a time when almost everyone will have one. Smart phones and other personal devices have embedded GPS capabilities and serve as the platform for many types of detection and image capture applications. Personal and business transactions are increasingly being performed on smart phones, and many types of applications are being redesigned to accept input from these mobile platforms, including the Next Generation 911 emergency response system. The emergency response community is actively looking for ways to tap the power of smartphones to support disaster preparedness and response operations, where interchange between responders and the public is critical. In a similar way, new generations of wearable devices can be used to closely monitor personal health status and may one day communicate directly with autonomous vehicles to improve pedestrian safety.
  We are moving towards a future where the citizen will become a major contributor of accurate, real-time information that will improve the quality of life of everyone.
- **Social media**: a variety of new application platforms, including Facebook, LinkedIn, Instagram, Twitter, and many others, enable individuals to join together for a variety of personal, recreational, community, political, and job-related interactions and activities. The billions of people who participate in these networks inevitably reveal enormous amounts of information about themselves: their personal characteristics, thoughts, and activities, often tied to their location. Data scientists have learned that valuable intelligence, for both good and bad purposes, can be derived from this data. On the positive side, social media-related big data analytics can help governments to better identify and respond to social problems.
- **Remote sensing technologies**: new generations of cheaper and more powerful large-scale data collection sensors, mounted on satellites, aircraft, and terrestrial vehicles, are making it possible to detect characteristics of our natural and built environment to a degree never before possible. Light detection and ranging (lidar), ground penetrating radars (GPR), synthetic aperture radars (SAR), hyperspectral, thermal, and magnetic field sensors are able to penetrate the ground and see through walls to identify important features and characteristics. New advances and applications for lidar should be especially noted. They have become invaluable to mapping the elevation of land and the exterior of structures. They are now also being used to collect information about building and tunnel interiors. In coming years, sensor technology should advance to the point where it sees beneath the street pavement to detect the location and depth of underground infrastructure.
- **Fixed sensors and the Internet of Things**: sensors are now being attached to, or incorporated into, everything from household appliances to streetlight and signal poles, and utility valves, pipes and house connections. High-capacity sensors attached to cars and trucks will provide the foundation for autonomous vehicles. Sensors capture information about the operations of the things they are connected to and also give them a precise location. Sensors attached to street furniture can monitor air quality in real time. Vehicles' sensors can alert authorities to traffic jams and accidents as they occur. Sensors and smart valves embedded within underground infrastructure networks can monitor utility performance, provide alerts about hazardous underground conditions, and customize utility services to fit the needs of each household.
- **Video capture**: Fixed and mobile video sources are increasingly being used to improve security. In cities like London and New York, video devices attached to street

31

furniture, transit stations, and buildings work as part of an interconnected security system. Video device location and the spatial identity given to each pixel in a video image are essential for efficiently searching and relating multiple video streams, each generating enormous volumes of data. Pattern recognition techniques are now starting to be employed to rapidly search video images to identify dangerous behavior that might require a quick response.

31.3.2 Major Spatial Technology Platforms

Smart Cities rely upon ever more effective applications of information. We are living in an age where, as described above, vast amounts of new, spatially enabled information are becoming available. This data tsunami flows from four major overlapping technology platforms all of which now have the ability to create and share spatially enabled data:

- **Integrated, spatially enabled enterprise platforms** that create a spatial information foundation within an urban area enabling government, businesses, and other organizations to manage their operations more successfully. This platform also includes operating and transactional data created by spatial enterprise users of all kinds.
- **Personal technology platforms** such as home computers, smart phones, and wearables that can be used to both actively and passively create and share data through a wide variety of transactions. Social media data is created on this platform and is connected together by the internet.
- **Remote sensing platforms** that collect location data about the physical and environmental characteristics of the Earth, often in broad sweeps from satellite or fixed wing aircraft, and local sensor data that monitors a narrow range of inputs from a fixed or mobile vantage point.
- **The Internet of Things (IoT)** that provides information about, and the location of, everything from underground utility valves to household appliances, giving private citizens and business managers greater information about and control over their assets.

Each of these platforms increasingly generates data of greater detail, accuracy, completeness, and interoperability, for business applications, decision support, and analytics. Datasets, regardless of the platform of origin, can be organized and brought together by using their common spatial links, in any combination found valuable. Whether from a single "big" dataset generated from social media or a combination of many discrete datasets from local government, this expanding universe of spatial data can be acted upon by the special spatial capabilities described in Sect. 31.2 of this chapter to create new levels of smartness. In short, as the volume of spatially enabled data grows, the value of spatial systems will continue to expand.

We have only begun to figure out how to best utilize our burgeoning spatial data capabilities. The data science programs of many colleges and universities, private research centers, government developers, and entrepreneurs are all hard at work looking for data combinations, analytic techniques, and technologies that produce effective results. Section 31.4 will focus on the key support elements that must be present in order to facilitate the full exploitation of spatial capabilities.

31.4 Spatial Standards, Data Models, Architecture, and Organization

31.4.1 The Role of Software Engineering in Smart Cities

To this point in the chapter we have underscored the importance to Smart Cities of establishing an enterprise GIS to give standardized spatial identity to everyone and everything to be found within an urban area. We have gone on to identify special spatial capabilities that extend the analytic powers of spatial systems beyond conventional IT applications. We have also described the characteristics of a greatly expanding data environment where spatially enabled data interoperability supports more effective applications and more powerful analytics.

The job of making sure that there is an effective spatial data infrastructure to serve as the platform for these new capabilities and applications is a challenging one. It requires that spatial data standards be designed and implemented across information generating platforms, so that spatial data fields can be confidently used to link datasets together. Also important are standardized data models that enable even tighter integration of the datasets within specific use categories such as underground infrastructure (UGI) and Building Information Modeling (BIM) systems. Furthermore, to enable awareness, access, transport, and analysis of spatially enabled data, a technical architecture must be put in place to handle the high-volume flows of information needed to support demanding uses, including big data analytics and artificial intelligence applications requiring massive flows of real-time data. Finally, knowledgeable managers and technical experts must be in place, properly organized and fully empowered, to ensure that the full value of spatially enabled systems can be realized.

31.4.2 Data Standards

The reason computer and telecommunications technology exist at all is due to the value of data, especially when that

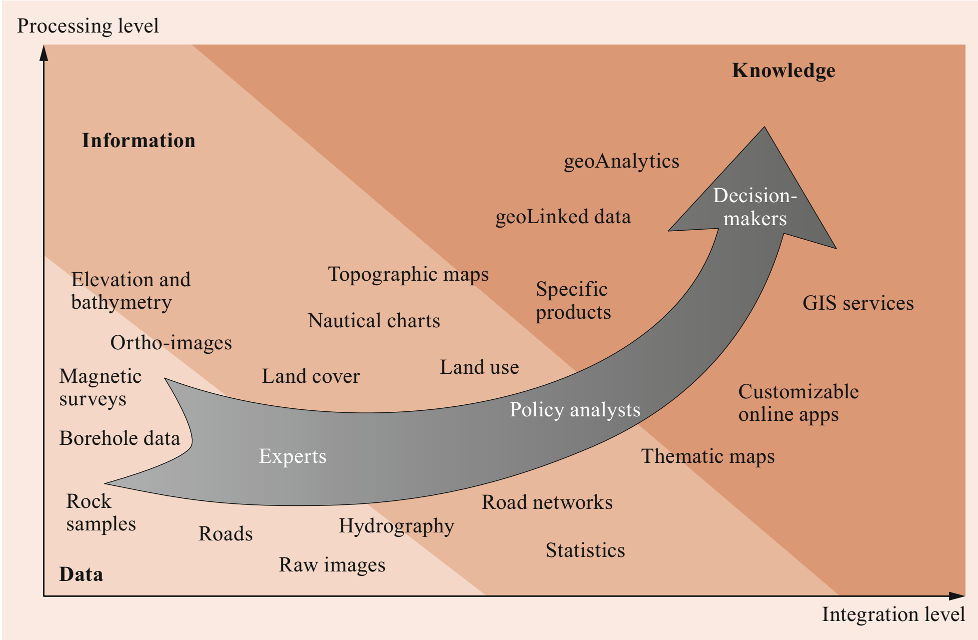


Fig. 31.6 A continuum of data to knowledge. After [19]

data can be transformed into operational intelligence. Raw data, with all its inevitable flaws and shortcomings, is not good enough, especially if it needs to be used for life-saving missions and the support of vital services for millions of people. In Sect. 31.2, we spoke about the unique capabilities of spatial data; however, these capabilities cannot be utilized if data is incompatible and of poor quality. It is the job of spatial standards to define the characteristics of data that will make it accurate, complete, and interoperable. Figure 31.6 illustrates how increasing levels of data integration and processing capacity enable applications of increasing complexity and value. Without standards there would be no assurance that data was trustworthy of being used effectively together.

A discussion of spatial data standards must begin with the work of the Open Geospatial Consortium (OGC).

To quote from their website,

> "The OGC is an international consortium of more than 500 businesses, government agencies, research organizations, and universities driven to make geospatial (location) information and services FAIR – Findable, Accessible, Interoperable, and Reusable.
>
> OGC's member-driven consensus process creates royalty free, publicly available, open geospatial standards [20]. Existing at the cutting edge, OGC actively analyzes and anticipates emerging tech trends, and runs an agile, collaborative Research and Development (R&D) lab – the OGC Innovation Program – that builds and tests innovative prototype solutions to members' use cases" [21].

OGC was founded in 1994 and has already passed its 25-th anniversary. The OGC Technical Committee is responsible for standards development and works through a series of subcommittees, domain working groups, and standards working groups. Among OGC achievements is the consensus adoption of KML (formerly Keyhole Markup Language), a widely accepted method for visualizing geographic data on a virtual globe:

> "Geographic data adds tremendous value to the online experience. More and more people are looking for ways to incorporate location information into their online content," said Michael Weiss-Malik, KML product manager for Google. "The standardization of KML makes it possible for both novice and expert users alike to publish and share geographical information in an open format. It's not unlike web browsers' standardized support for HTML, which allows any web browser to read any web page" [22].

The OGC Web Services (OWS) suite of standards [23] provided the needed consensus to achieve the first generation of spatial data infrastructures. Through endorsement of OWS standards by GSDI, several national SDIs, and the trans-European SDI known as INSPIRE, thousands of web services are now offering hundreds of thousands of spatial data layers for access by anyone with a web browser. The OWS suite consisting of the Web Map Service (WMS), Web Map Tile Service (WMTS), Web Feature Service (WFS), Web Coverage Service (WCS), and Web Processing Service (WPS) continues to evolve to the latest web technologies. A draft version 3 of WFS—refactored to a resource-oriented design and utilizing the latest API technologies—has been released in 2019 for testing and public comment, and can be found at OGC Web Feature Service 3.0: Part 1 – Core (opengeospatial.org).

Building on the operational OGC Sensor Web Enablement standards, OGC's SensorThings standard enables fixed and mobile sensors of all types to "talk" to each other.

> "The OGC SensorThings API provides an open, geospatial-enabled and unified way to interconnect the Internet of Things

(IoT) devices, data, and applications over the Web. At a high level the OGC SensorThings API provides two main functionalities and each function is handled by a part. The two parts are the Sensing part and the Tasking part. The Sensing part provides a standard way to manage and retrieve observations and metadata from heterogeneous IoT sensor systems. The Tasking part is planned as a future work activity and will be defined in a separate document as the Part II of the SensorThings API" [24].

31.4.3 Standardized Data Models for Key Smart City Use Cases

OGC looks to identify important business use cases where data silos and a lack of standards inhibits data interoperability and undermines efficiency and effectiveness. This includes initiatives related to the 3-D urban landscape (CityGML), Building Information Modeling (BIM), underground infrastructure (UGI), and emergency preparedness and response. OGC utilizes a standard methodology to approach these thematic areas.

CityGML

CityGML is an OGC-developed open data model and XML-based format for the storage and exchange of virtual 3-D city models. It is an application schema for the Geography Markup Language version 3.1.1 (GML3), the extendible international standard for spatial data exchange issued by the Open Geospatial Consortium (OGC) and ISO/TC 211. The aim of the development of CityGML is to reach a common definition of the basic entities, attributes, and relations of a 3-D city model. This is especially important with respect to the cost-effective sustainable maintenance of 3-D city models, allowing the reuse of the same data in different application fields [25].

Building Information Modeling (BIM)

The most efficient and effective way to manage and maintain a building is to have integrated information about all aspects of a building's physical foundational, structural, and mechanical components, including its connection to the ground and to the utilities that provide it with water, power, and other services. This requires that data about building design, construction, and maintenance be integrated into systems that manage and optimize all aspects of building operations. Sensors, videos, and interoperable digital drawings and maps must be combined through all phases of design, construction, and maintenance processes. Since a building is a vertical map made up of interconnected utility and structural networks and systems, BIM is heavily supported by spatial data and spatial analytics [26].

Underground Infrastructure (UGI)

The subterranean area beneath city streets and buildings has always been terra incognita. Until only recently information about water, sewer, gas, electric, telecommunications, steam, and transit networks was locked away in utility data silos, difficult to access even by the data owners. However, there is now increasing recognition that rapid access to accurate, standards-based utility information can support a wide variety of use cases, including excavations, asset management and emergency and disaster response (Sect. 31.5.6). Utility data becomes much more valuable when the data from any one utility is interoperable with the data from other utilities, creating new analytic options and enabling collaborative work between utilities. In a disaster event, rapid access to interoperable infrastructure information is indispensable and can be a life saver [27].

31.4.4 Key Architectural Elements

Figure 31.7 presents a framework that depicts all the key elements of a spatial data and applications architecture. This architecture will only work effectively if the spatial data it supports conforms to standards and is, therefore, interoperable. Several key components of this architecture are noted below:

- **The IT foundation**: many business and government organizations have established their computer operations on a conventional IT foundation. Because the development of spatial information systems lagged behind the development of IT, most computer infrastructures were not designed to support the larger volumes of spatial data, nor the need to bring spatial data together from many different sources. As the size of spatially enabled datasets continues to grow, current IT infrastructures will need to be redesigned, and technicians and managers with a solid spatial understanding will need to be hired.
- **The cloud**: the creation of massive computing environments with enormous storage, processing, and telecommunications capacities is perfectly suited for the spatial data operations of the future. Instead of individual organizations creating high-capacity, customized environments of their own, they can use cloud resources, almost always at much lower cost. At times of extremely high usage, such as during an emergency or disaster event, local servers are almost always overwhelmed, while a cloud environment has the potential to provide almost unlimited amounts of additional capacity for as long as it is needed.
- **Security**: standards-based, spatially enabled data shared across organizations has the potential to solve problems and create new kinds of valuable intelligence. This requires that participating organizations believe their data assets will be safe from theft, tampering, and other destructive activities. Effective security is, therefore, an essential component of a successful spatial data architecture.

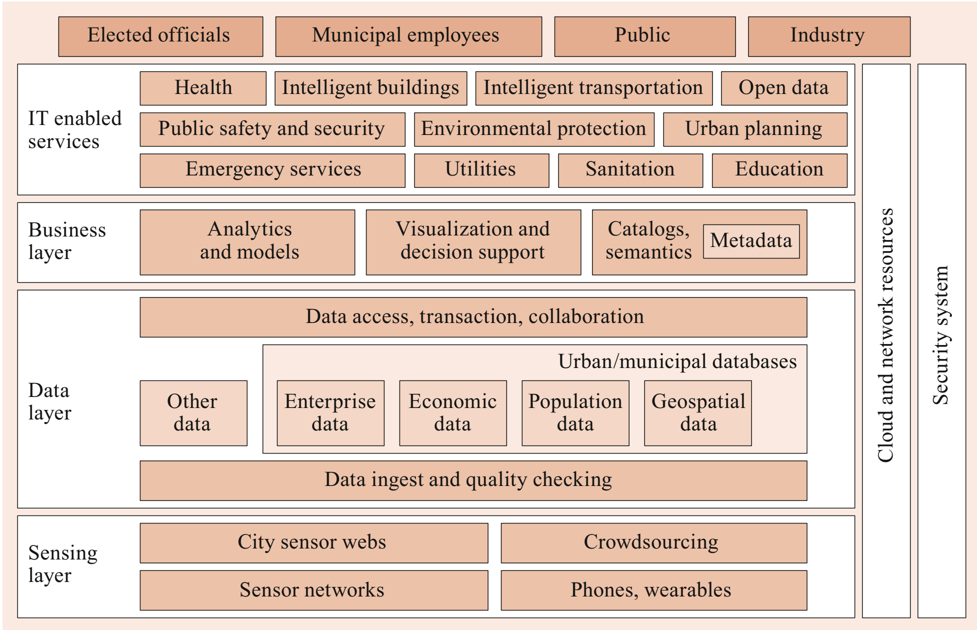


Fig. 31.7 Smart City ICT Framework, Enterprise Components. After [28]

- **Data discoverability, accessibility, and usability**: the increasing power of spatial systems cannot be reserved for use by technicians alone. While there will always be areas that only highly trained experts can master, it is important that spatial infrastructure be designed to enable access and use by anyone. This will require the building of easy-to-use spatial data search tools that can discover datasets of interest, simplify their download and integration, and enable analytics through dashboards and map-based graphic user interfaces.

31.4.5 The Spatial Organizational Infrastructure and Policy

An enterprise computer environment that is heavily based on spatial data and analytics requires a personnel structure that differs from the traditional IT organization chart. Adherence to old hierarchies and obsolete cultures will not produce the kinds of benefits that spatial capabilities have the potential to provide. Spatially trained personnel are among our most valuable resources. They must be provided with an organizational framework that empowers their skills and brings out their creativity. The following are a few examples of the kinds of organizational practices that need to be adopted:

- **Chief Geospatial Information Officer (CGIO)**: many traditional IT shops either fail to have a spatial division or bury their spatial expertise beneath layers of bureaucracy, cutting them off from decision-making, essential staffing, and necessary funding. This is tantamount to reducing the potential effectiveness of IT operations by half, if not more. It is essential that IT organizations recognize the importance of spatial capabilities by ensuring that the CIO has a CGIO directly reporting to her and is provided with staff and funding to do her job properly. An organization's GIO must be more than a good technician. She must be able to convince decision makers about the value of spatial capabilities, and must effectively manage budgeting, procurement and maintenance processes. The CGIO must be directly involved in all major IT decision-making, because the effectiveness of spatial systems is contingent on the suitable design of the underlying computing infrastructure and on the overall IT strategy.
- **Spatial steering committee**: a sizable city could well contain 20 or more agencies, each with their own GIS division, each division managing a number of spatial databases and applications. At the same time, each agency will depend upon the creation and maintenance of foundation spatial datasets like imagery, streets, buildings, parcels, and addresses. To achieve the most effective management and operation of a city's spatial infrastructure, a steering committee of agency spatial managers needs to be formed. Such a committee will support data sharing, collaboration on common applications, investment in spatial data and technology infrastructures, and the evolution of an effective architecture to support agency and citywide missions.
- **Spatial strategic planning**: given the importance and power of spatial systems and recognizing that spatial capabilities must be coordinated across all sectors of a jurisdiction and of a region or state, it is essential that CGIO's be given the authority and resources to develop a spatial strategic plan that is integrated with the IT strategic plan but that extends beyond the boundaries of a specific agency to embrace the jurisdiction as a whole. The spatial strategic plan must also be integrated with

smart city policies and programs, because almost every smart city initiative will have a spatial component.

- **Innovation through prototyping**: modern IT development relies on rapid experimentation to identify novel methods to extend existing capabilities. A geospatial innovation culture regularly conducts hackathons and other initiatives that focus on software developer interactions. Low-cost, open-participation iterative experiments, testbeds, and pilot projects based on free and open interface and encoding standards encourage innovation and provide insight and guidance that can optimize for improvisation and resilience, as well as prevent expensive IT failures.
- **Architecture and policy**: a key objective for the management of smart city information technology and its benefits is to convert short-term successes into enduring value in municipal government. Vendors are at the forefront of developing new spatial capabilities for cities, states, and counties. However, many of these innovations are dependent upon a single provider due to a lack of common interfaces. Software engineering practices oriented towards open interfaces and encodings that support interoperability are critical. The more advanced smart city programs are designing robust software architectures that incorporate open standards, thereby guaranteeing greater data interoperability and the easier sharing of common tools and applications.

The OGC Smart City Interoperability Reference Architecture (SCIRA) project [29] is an example of the technical structure and policies needed to support spatial smart cities. The SCIRA project is researching, designing, and testing a reference architecture as an interoperable framework that integrates spatial software, IoT sensors, and municipal IT for public safety applications at the community level. The SCIRA project is developing an requirements-based architecture framework and deployment guides to support development and interoperability for Smart Cities. The deployment guides will make the highly technical architecture accessible and useful to procurement activities. The structure of the SCIRA development is based on the following methodology:

1. The City Manager establishes a deployment strategy.
2. Based on the strategy, City IT Managers work with their procurement departments to analyze deployment needs.
3. From those deployment needs, the City IT Managers identify deployment requirements (typically as part of a broader set of system requirements).
4. A logical architecture is then defined, based on the SCIRA reference architecture and using a standard architectural framework such as RM-ODP, TOGAF, etc.
5. During procurement, Commercial Services Providers submit their interpretations of the deployment architecture (e.g., UML deployment diagrams) based on the logical architecture that was defined by City IT Managers.
6. The selected Commercial Services Providers will then implement, test, and deliver the system.

SCIRA development guides will support evolution of the IT baseline, including multiple procurements over time for similar components in the architecture. The architecture, including the use of open standards, avoids vendor lock-in that limit city options.

31.5 Use Categories and Use Cases for Spatially Smart Cities

31.5.1 Delivering Value

Looking back over the past 20 years it is possible to identify a large number of use categories and use cases that depend upon or are improved by spatial capabilities. In many instances, it is possible to speak about not merely spatially enabled applications but spatially *dominant* applications, where operational effectiveness would not be possible without spatial components. Presented below are ten such use categories, each of which includes descriptions of several use cases. In every instance, data is combined from multiple sources and is subjected to a variety of spatial operations and analytics to support applications with Smart City characteristics. Through these examples we can understand just how pervasively spatially enabled systems have been deployed across those municipal governments that have developed an underlying spatial data infrastructure. It should be noted that in such jurisdictions, spatially enabled applications often constitute the primary form of IT deployment. In these examples, we also see strong intimations about a future where more powerful spatially enabled systems and technologies continue to extend the power of IT and make bigger and smarter impacts on our society. Many examples are drawn from one writer's NYC background. However, NYC's spatial efforts have mirrored, and often been derived from, similar efforts by local governments across the US.

31.5.2 Citizen Engagement

- **Background:** Citizen engagement in ways that enhance quality of life and support democratic principles is a foundational characteristic of smart cities. While cities have traditionally interacted with and served their citizens, new spatial tools are enabling even greater levels of collaboration because communications concerning an issue or problem is meaningless unless the information imparted has a location identity.

- **Smart Spatial Approaches:**
  - **Citizen interaction with municipal services:** cities provide a wide range of location-based services on a day-by-day basis including garbage collection, street cleaning, school bus routing, fire and police protection, and building, health, and restaurant inspections. When municipal government is operating at its best, service delivery personnel and inspectors from dozens of agencies regularly communicate with citizens and businesses, responding promptly and efficiently to needs, while collecting information that will determine how services can be improved.
  - **Open data and public participation:** many municipalities and states are starting to put their spatial base layers and location enabled service information into open data portals for citizen access. This enables individuals, groups, and businesses to access vast amounts of data that can be used to evaluate services and assess the adequacy of proposed initiatives. Using smartphones and computers, citizens can rival government in their ability to understand what is going on in their communities, and to provide more effective feedback to elected and agency officials.
  - **Citizen as collaborator:** municipal residents can be an invaluable source of information for municipal officials providing feedback about conditions in communities in real time. 311 systems dispatch agency work crews in response to service requests and complaints. Now, new social media tools provide government with instant feedback about policies and about neighborhood and household conditions. During emergencies and disasters these communications tools, aided by precise location, can assist responders to protect and rescue citizens caught within a disaster area.
  - **Customization of services:** the evolutions of new spatial technologies, sensors, and geo-enabled data are leading to new kinds of services customized to the needs of individuals. Citizens can use a smartphone to summon a vehicle that will take them from wherever they are to wherever they wish to go. Other applications provide a wide variety of pickup and delivery services. These applications invariably use spatial capabilities such as real-time GPS and vehicle routing.
- **Use Case:** NYC Open Data Crime Map: this writer remembers in the mid 1980s when as president of a Brooklyn block association, he attempted to get crime statistics for his neighborhood, because violent incidents had become an almost daily occurrence. It was impossible to get a straight answer from the local precinct even though the police kept excellent crime records by location. While he was told that his area was the low crime end of the precinct, he eventually learned, through unofficial channels, that in fact, out of eight sectors, the one he lived in had the second highest number of recorded crimes. This can no longer happen in New York City. Based on a 2012 law, in 2013 NYPD began making available to the public an online service called NYCCrimemap, which makes available crime statistics and crime maps [30]. This information is helping to transform police work in New York City. Citizens are better informed about the risks they face in their neighborhoods, and they are more inclined to assist police officers patrolling in their community. This has fostered closer relations between the police and the communities they serve. Public crime information plus CompStat and Real Time Crime programs, all spatially enabled applications, have been among the factors that have led to a decline in murders from 2245 in 1990 to 292 in 2017, a reduction of 87%.
- **Future Prospects:** The open data movement, which makes spatially enabled information about the city and its operations available to the public, is starting to gain traction. Its use is held back by the challenges presented by spatial datasets that have not been standardized for easy interoperability and by open data portals that require users to have advanced technical skills. We look forward to a time when the spatially enabled data in open data portals is made fully interoperable and accessible using intuitive dashboards and preprogrammed analytics.

31.5.3 Property Assessment and Taxation

- **Background:** Property taxes are an essential revenue source supporting municipal operations and local boards of education. An average of 30% of municipal revenues come from property taxes [31]. Property tax bills are based upon property value that is determined by municipal assessors. The assessment process was originally paper based, and all work processes were manual. Over the past 40 years almost all of this work has been computerized. For example, parcel maps have been digitized and are used with data layers providing detailed descriptions of residential, commercial, and industrial buildings, and other structures. Computer assisted mass appraisal [32] (CAMA) systems store and process assessment data and computerize tax billing and collection processes. Property tax assessment and collections systems are among the most spatially dominant of all municipal applications.
- **Smart Spatial Approach:** In recent years, new data sources have been developed that support the assessment process, allowing it to be more efficient and accurate. High-resolution aerial photography overlaid with building footprints and parcel maps allows appraisers to examine the backyards and rooftops of properties, the parts of properties that are generally hidden from the view of assessors on the street. 360° oblique angle photography

showing all building facades, when integrated with overhead imagery allowed for an even richer imagery context. In addition, combinations of data from building inspections, construction permits, violations, rental rates, and resale value, all keyed to building address, parcel number, and building identification number, allow assessors to get a more complete picture of a property's worth. Municipalities report a 2 to 5% increase in revenues as a benefit of using these new, spatial data driven methods [33].

- **Use case: NYC Department of Finance Tax Assessment:** The New York City Department of Finance, through a sensor-based data collection contract with CycloMedia Inc., has recently started to incorporate street-level photography and lidar into its assessment processes. Assessors are now examining properties in 3-D, enabling more accurate quantification of floors and floor area. The new data also provides clearer evidence of property improvements and extensions that might increase values. Assessor productivity has risen because fewer visits to the field are necessary since combinations of imagery and data layers viewed on a computer screen produce a data environment often superior to observations from the field. Due to the incorporation of street-level lidar alone, property tax billing may increase by 4%, providing as much as a billion dollars in additional revenue.
- **Future Prospects:** Beyond current methods there is even greater potential for increased accuracy and completeness in the assessment process. Lidar technology capable of capturing details about building interiors, floor by floor can add valuable new data for assessors to consider. Artificial intelligence tools can help assessors to manage the vast array of data to produce more accurate assessments. Assessors are starting to extend their spatial analysis to include unused air rights above built structures and to underground easements. As an added bonus, comprehensive imagery of building interiors and exteriors can be valuable to emergency responders.

31.5.4 Customer Relationship Management (311 CRM Systems)

- **Background:** Most cities and counties across the US offer an array of public services that include management of water and sewer services, roadway maintenance such as pothole repair, building inspections, restaurant inspections, street cleaning, garbage collection, and many more. Traditionally, each agency developed its own methods and systems for handling service requests, including complaint intake, routing to field offices, and dispatch of appropriate field units to appropriate work locations. These methods varied by degree of automation, accuracy, and completeness of information. Data for each of the different service categories was rarely interoperable, so analytics across services was difficult, if not impossible.
- **Smart Spatial Approach:** Spatially enabled systems are giving municipalities and counties the ability to improve their complaint handling and service delivery operations. The addresses of requestors are validated using a geocoding application, and the appropriate service yard is automatically notified. Work tasks are prioritized and routed so work crews can maximize effectiveness and productivity. Suspected duplicate calls are flagged so that workers are not dispatched multiple times for the same problem. Backend analytics, acting on data, standardized across yards, provides valuable intelligence for improving operations.
- **Case study: 311 CRM Systems:** Many counties and municipalities have taken their spatial approach one step further. They have now adopted integrated customer relationship management (CRM) systems that funnel complaints and service requests of every type through a 311answer center. In this way, the public needs to know only one three-digit number to enter a request for any one of dozens of different municipal services. CRM introduces more efficient handling of service and complaint requests, standardizes the information recorded, allows citizens to access complaint status information, and vastly increases the types of analytics that can be performed, such as the spatial distribution of requests by time of day. Relationships between different complaint types can now be examined. 311-related communications now enter the system via landline phone, smart phone, and email. Service data is increasingly being provided to the public and community groups along with tools to perform analysis of the information.
- **Future Prospects:** The future for CRM systems is promising as capabilities expand. Easy to use map-oriented dashboards are being provided, allowing citizens to track their complaints more easily, and enabling community groups to closely examine service delivery effectiveness in their neighborhoods. Structured crowdsourcing of information can now provide city service agencies with real-time videos and photos of problems. Increasingly, sensors and predictive analytics will allow agencies to get out ahead of certain kinds of complaints and deal with issues before they become problems. In time, the complaint systems of private utilities will be integrated into those of government agencies creating the potential for even greater efficiencies and improved collaboration between different field response teams.

31.5.5 Public Safety

- **Background:** Across the US police, fire, and emergency (EMS) services are responsible for saving hundreds of lives daily. Key to reducing deaths and decreasing the severity of injuries is the time taken by trained responders to arrive at the scene of an incident. Effective response is highly dependent upon rapid notification of the local 911 emergency call center (Public Safety Answering Point or PSAP) including providing information about incident type and location. PSAPs then identify and dispatch the best situated field response unit to the scene. Google estimates that as many as 10,000 lives can be saved for each reduction of 1 min in response time [34].

- **Smart Spatial Approaches and Use Cases:** The uses of location are being enhanced by advancements in 911 dispatch systems and also through a number of newer applications that are combining data sources, advanced sensor information, and more powerful data analytics. Use cases are the following.
  - **Next Generation 911:** The public safety community, through the efforts of the National Emergency Number Association (NENA), is developing Next Generation 911 (NG 911) [35], designed to accommodate smart phones and other mobile devices, which are replacing landline phones as the primary means of communicating emergency information. NG 911 can use the GPS capabilities of smart phones to rapidly pinpoint the location of an incident and can also incorporate location-tagged video and photos sent from the scene to improve awareness and to serve as possible evidence in cases of crime or negligence.
  - **CompStat** [36]**:** Police analytics are based largely on mapping the location of crime incidences, by type, method and time, to identify crime patterns. The intelligence derived from these analytics supports the work of investigators and also helps police to improve their tactics. CompStat, considered a primary example of intelligence led policing, has been shown to be an important factor in reducing crime rates.
  - **Real Time Crime** [37]**:** Police departments in large cities such as Miami, Seattle, Houston, and New York have set up information centers to support police response to crimes in progress. Police in these cities can contact their Real Time Crime Center (RTCC) and rapidly obtain information about individuals with criminal records and the history of crime incidents in the immediate area of response activities.
  - **Gunshot location:** Police departments are deploying arrays of gunshot sensors in neighborhoods subject to violence, where underreporting by the public is an issue. The sound of a gunshot is picked up by multiple sensors networked to a central computer running spatial analytics that can accurately identify the gunshot's point of origin. This provides responding officers with a rapid heads up about where to direct their investigations and provides a measure of safety through improved situational awareness.
  - **Video surveillance systems:** Networks of video cameras are being strategically deployed in sensitive and high crime areas to collect footage that can be used in criminal investigations. Depending on the location of the criminal activity, video recordings are searched to identify suspects and track their movements. These systems have been successfully deployed in London, New York, and other major cities.
- **Future Prospects:** New data sources, improved technologies, and advanced analytics will continue to improve emergency services in the future. Smartphones and wearables can be designed to monitor the health of individuals and to send a distress signal to responders containing precise location information. Data from a variety of sensors and cameras, combined with real-time data searches, will give police and fire personnel comprehensive information about incidents in progress, enabling safer and more effective responses. Sensors embedded in the smart phones carried by individuals or located in fixed positions in public and private places can be used to detect hazardous chemicals, pathogens, and explosive materials, before they become a threat. Applications are being developed to allow emergency vehicles to control traffic signals and to weave through traffic more easily, enabling them to reduce response time.

31.5.6 Disaster and Emergency Preparedness and Response

- **Background:** Large-scale disasters and smaller-scale emergencies are spatial events. They occur at specific locations and affect people, structures, and infrastructure, all of which have location identities. Responses to these events involve rescuing people from somewhere and sending them to safety somewhere else, and transporting responders and resources from outside the impact zone to locations inside. In the past, disasters struck unexpectedly and rescue and response efforts occurred in conditions that resembled an information blackout due to breakdowns in communications. Efforts to mount an effective response are often overwhelmed by the enormous number of responders, whose activities need to be coordinated, and the huge volumes of data that pour into the emergency operations center. Now, increasingly sophisticated spatial information support coupled with improved communications is allowing progress to be made.
- **Smart Spatial Approach:** Preparedness – the deployment of spatially enabled enterprise systems containing interoperable data layers gives emergency planners the ability to model the potential effects of an event. For example: information about critical infrastructure allows planners to identify vulnerabilities, single points of fail-

ure, and triggers to cascading effects. Pre-events spatial analytics enables planners to selectively harden key facilities and to develop strategies to keep people safe. Prediction: spatially enabled remote sensing technologies and systems allow emergency managers to anticipate events by tracking storms, predicting rain volumes and wind speed, and by sensing the dryness of woodlands, the shifting of underground faults, and many other environmental and geological factors. Response: the deployment of spatially enabled mobile technologies in the hands of first responders allows rapid determination of damage on a block-by-block, structure-by-structure basis, while wireless communications allows the microview of each responder to be assembled centrally into a comprehensive map that gives incident commanders the real-time-information they need to make decisions. Smart phones in the hands of citizens caught within impacted areas gives them a tool to communicate their situation and their exact location to emergency responders. The logistics of moving responders and resources into the disaster or emergency area is being spatially coordinated.

- **Case Study – The Spatial Response to the Attack on the World Trade Center in NYC:** Although 9/11 occurred more than 20 years ago, the response of the Emergency Mapping and Data Center (EMDC) established by NYC OEM demonstrated the importance of spatial smart city capabilities in dealing with the chaos of a major disaster. In 1999, NYC had completed the development of its planimetric basemap; composed of imagery, streets, and structures; and was actively building a comprehensive repository of data layers registered to those foundation layers. This data had been distributed widely across city agencies. Within minutes of the collapse of the Twin Towers, the Phoenix Unit of the Fire Department (FDNY) began printing maps to guide rescuers on the scene. The Phoenix Unit working with FEMA (Federal Emergency Management Agency) also developed a map grid to organize search and rescue efforts, and to delineate areas already searched and hazards to be avoided. On the afternoon of September 11th, satellite imagery was used to capture the first comprehensive overhead imagery of the devastated site, which was immediately put to use by city leaders. Imagery, lidar and thermal, and multispectral imagery, whether from satellites, fixed-wing aircraft, or fire department personnel hanging off the landing skids of a police helicopter, was captured on a daily basis to estimate the volume of debris, detect chemical hazards, and locate hotspots. Information about the disaster area was communicated to the public using interactive mapping applications, letting people know which subway stations and tunnels were open for use. The remains of victims were pinpointed by location coordinates obtained from GPS devices and entered into a spatial application loaded into newly deployed mobile devices. Parcel and structure information was used by inspectors from the Buildings Department, also supported by location-enabled mobile devices, to evaluate structural damage in a wide area around the immediate impact zone. Private utility companies and infrastructure agencies were asked to gather their information for delivery to the EMDC, where it was scaled and oriented to the City's basemap and transformed so that all the layers could be viewed together. Amazingly, all this information, because it was related to a common spatial foundation, was interoperable. Because the spatial response to 9/11 utilized the city's enterprise GIS and accessed and integrated so many different kinds of data in support of dozens of response activities involving thousands of responders, we believe it embodied many of the characteristics associated with Smart Cities and judge it an appropriate beginning point for the concept of spatially enabled Smart Cities.
- **Future Prospects:** Since 9/11 the NYC Department of Emergency Management along with the Fire Department and the Police Department have improved their spatial capabilities, but there remain a number of steps that still need to be taken. Extreme disaster and emergency scenarios need to be simulated in advance to better identify vulnerabilities and design sensible mitigating actions. Better spatial data exchange methods need to be implemented to improve field communications and inter-agency collaboration moving towards the design of a true common operating picture (COP). Machine learning techniques need to be devised to automatically organize, prioritize, and analyze the enormous volumes of information that flood into Emergency Operations Centers and would otherwise overwhelm responders. Data products, designed in advance and customized by location for specialized field teams, need to be developed. Also, improved strategies must be devised to leverage spatially enabled social media data, also known as crowdsourcing, in ways that improve communications between the victims trapped by a disaster and the response community to facilitate faster and more effective rescue operations.

Emergency response was a major focus for early smart city developments, and a number of initiatives were led by large vendor organizations. Command and control information technology was relatively easy to deploy in the highly structured environments of emergency management. Smart City technology designed by IBM became a showcase in the Operations Center of the City of Rio de Janeiro, Brazil [38]. The centralized center was effective in achieving its defined mission. Broadening centralized centers to include SDI concepts will bring further capabilities for safe Smart Cities.

31.5.7 Underground Infrastructure Management

- **Background:** Underground infrastructure, including networks of pipes, conduits and tunnels for water, gas, sewer, steam, electricity, telecommunications, and transit systems are the critical underpinning of neighborhoods and urban centers. A picture of the underground environment would not be complete without including underground soils, streams and water table, and other geological features that comprise the physical environment through which utility networks run. In the 1970s and 1980s manual engineering methods gave way to computer-aided design and drafting (CADD) systems, and then, more advanced municipalities and utilities were able to combine their CADD drawings into continuous, fully networked GIS layers that facilitated capital planning, system-wide modeling, and a graphical interface for maintenance and management systems. These GIS utility layers often had several drawbacks: Failure to register to a common, accurate basemap meant the locations identified for utility features were not spatially accurate and could not be relied upon when excavating. Failure to have a common spatial reference standard meant that different utility layers could not be overlaid on each other and used together. It also meant that different suppliers of the same utility services operating in adjacent jurisdictions could not seamlessly join the representations of their networks for regional modeling, planning, and joint operations. Shortcomings in GIS software limited their ability to represent utility layers in 3-D, creating great uncertainty during excavations about utility depth. Although CADD and early GIS utility network mapping represented a leap forward for their time, they had significant limitations.
- **Smart Spatial Approaches:** The Open Geospatial Consortium (OGC), the international organization responsible for spatial data standard setting, initiated an underground infrastructure data interoperability project (UGI Project) [27] to address the problem of incomplete and incompatible underground information. In addition to OGC, UGI Project sponsors include the Ordnance Survey of Great Britain, the Fund for the City of New York, The City of New York, and the Singapore Land Authority. An initial project task was to better understand how the inadequacies of current underground information result in inefficiencies. The project team determined that poor data on utility placement and the underground environment led to the following problems.
  - The expense, delays and inaccuracies associated with physically marking the location of utilities on the street surface, in advance of excavation
  - Accidental utility strikes due to inadequate utility location information
  - Construction delays due to unexpected utility interferences
  - The inability to assess underground dangers due to lack of knowledge about utility interactions with surrounding soils, resulting in premature wear, leaks, breaks, and loss of service
  - Delayed response due to lack of reliable information at the scene of a utility emergency due to fear of causing additional harm out of ignorance
  - Increased risks during disaster events due to lack of awareness of utility single points of failure, triggers for cascading effects. and critical interdependencies.

  The first output of the OGC UGI project is the high-end MUDDI model (Model for Underground Data Definition and Integration), which creates a framework for more detailed data feature interoperability that will allow key data from multiple utilities to be brought into a common computer environment that can be extended to field crews.
- **Case Studies – Flanders and Auckland:** The OGC UGI initiative has identified two jurisdictions that are leading the world in underground infrastructure data integration: Flanders, Belgium and Auckland, New Zealand. In Flanders, an accidental utility strike in 2004 led to a disastrous gas pipeline explosion that killed 24 people. In response, regional authorities mandated that all private and public utilities digitize their underground infrastructure data in 2-D to meet common data standards. Flanders' utility mapping system (KLIP) [39], created in response to the disaster, is now responsible for making available interoperable utility information for all excavations. The system provides comprehensive utility information within two days with faster response times being planned. The use of the KLIP system has substantially reduced utility strikes and construction delays.

  Auckland has mapped water, sewer, and geological layers in 3-D, which has enabled improved planning for their expanding population. Water and sewer infrastructure maps are connected to parcel, structure and demographic layers, allowing planners to model a variety of development scenarios. The Auckland spatial system also supports disaster preparedness since New Zealand has a history of severe earthquakes. Also of note: the Ordinance Survey of Great Britain is currently working on an underground infrastructure demonstration project in the City of Newcastle, while New York City's IT Department (DOITT) is conducting a utility mapping pilot project in Long Island City, Queens.
- **Future Prospects:** Once a comprehensive underground infrastructure data model has been developed and utility data quality has improved, the stage will be set for many smart applications. Access to accurate utility data in the field will speed excavation permits and support rapid response to utility emergencies. Analytics will predict

utility failure and prioritize inspection and replacement initiatives. Sensors and smart valves will continuously monitor infrastructure performance and alert maintenance personnel about problems. Smart building (BIM) connections to underground infrastructure will permit more customized and efficient utility services and will accommodate smart grid solutions. Smart stormwater systems can guide rooftop and roadway runoff through the sewer system and to green areas while reducing the load on wastewater treatment plants and reducing sewer overflows into natural water bodies. Additionally, connecting maps of underground infrastructure to BIM systems, and to distant transmission, generation and storage facilities, will enable regional planning and foster intergovernmental collaboration.

31.5.8 Health Planning and Disease Control

- **Background:** Dr. John Snow [40] is known as the father of modern epidemiology. In 1854, Dr. Snow used a simple paper map to understand the spread of cholera in a London neighborhood, eventually tracking the origin point of the disease outbreak to a polluted water well. Since then, maps have been used to reveal the pattern of disease incidence and spread, and to design methods for containment. With the advent of computer mapping and improved data gathering, and diagnostic techniques, health professionals working from the local level to the international level are now able to detect and track diseases, and develop control and eradication strategies, resulting in a large-scale reduction in mortality and disability.
- **Smart Spatial Approaches:** Health practitioners around the world are responsible for reporting the incidence of disease of all kinds by location, thereby making it possible to identify transmission pathways and local hot spots. This analysis is key to rapid and effective response that supports the eradication of diseases like polio and measles, and the control of the flu, Ebola, and tuberculosis. In some nations, real-time monitoring of sales records for certain over-the-counter drugs like antiviral medication, cough medicine, and analgesics, can signal disease presence, giving health professionals an early alert, thus, allowing rapid mobilization of resources. Spatial methods are also being used to understand chronic environmental health issues such as asthma, by analyzing incidence patterns with combinations of data relating to industrial pollution, vehicular emissions, the prevalence of cigarette smoking, the presence of pollen, mold and vermin; and climate conditions.
- **Case Study:** The West Nile Virus made its presence known in North America for the first time in the summer of 1999 in New York City, when significant numbers of residents fell ill with an unknown virus that proved fatal for vulnerable individuals. Eventually, the Federal Center for Disease Control (CDC) and the NYC Department of Health (DOH) diagnosed the disease as a mosquito-borne virus that was also associated with disease in select species of birds. The NYC DOH in collaboration with the CARSI Laboratory of Hunter College's Department of Geography [41] developed a multipronged approach to deal with the outbreak [42]. Using mosquito traps, and a public hotline to crowdsource the location of dead birds, the city was able to identify areas where the virus was active, often before human cases occurred! Areas testing positive for the disease were treated with insecticide. Public health workers also sprayed low-lying areas likely to have bodies of standing water, based upon an analysis of land use, elevation, and hydrographic map layers. In addition, the city used a digital map of all sewer catch basins, to support a program that organized the work of field teams, via routing algorithms, to seed each catch basin with insecticide pellets. The combination of these spatially dominant measures, brought the West Nile Virus under control, and NYC's methods have been adopted by other jurisdictions across the US.
- **Future Prospects:** Spatial systems will be increasingly used for early identification and suppression of disease outbreaks by putting increasingly effective predictive tools in the hands of health professionals. Sensors embedded in wearables and smart phones, and located in public places, may soon give public health officials an early alert that a particular pathogen is in active circulation. Sensors will also be used to better monitor disease in farm animals. Such knowledge can trigger special monitoring at airports and other mass transportation centers and along major vehicular routes, enabling infected but not yet symptomatic individuals to be identified and intercepted. In addition, sensors will be increasingly used to detect dangerous substances in the air, ground, and water supply, giving the public early warning, and enabling more effective preventive action.

31.5.9 Transportation

- **Background:** The ability to get from one place to another easily and quickly has always been a key driver of economic activity and urban development. With the advent of the internal combustion engine, roadway maps have enabled motorists to find their way. Applying street number sequences to the buildings along a street frontage became an essential method of achieving specificity of location, especially in cities where finding a particular building among thousands could be a huge challenge. The street network of a city can occupy more than 25% of urban land

and provides the rights-of-way for utilities. Spatial systems, ideally suited to map and model linear networks, are used for roadways, navigable waterways, and air traffic corridors; and can integrate information from a variety of transportation networks, to support comprehensive transportation planning and design.

- **Smart Spatial Approaches:** Keeping records of all aspects of a city's street system has been a basic function of municipal and county government from well before the digital era. Now, details of streets and mass transit are kept within spatial databases and are utilized for a wide variety of spatial applications and analytics. Spatial systems, tied into customer response management (CRM) applications, are being used to manage street light replacements, to fill potholes and to provide information in real time about which streets have been plowed during a snowstorm. Complex street signaling systems manage traffic flows and can be adjusted for the time of day. Mapping systems help to coordinate and ensure the safety of excavations – even if only via marking the location of underground utilities on the street surface. More recently, the use of smart phones with routing applications has revolutionized the way we get from one place to another and has spawned new customized car services like Uber and Lyft that match passengers with drivers based on real-time location and routing algorithms. The efficient routing of entire vehicle fleets responsible for work processes ranging from package delivery to waste disposal is now performed by spatial algorithms, enabling companies and government agencies to reduce vehicle mile and increase driver productivity.
- **Case Study:** NYC's Vision Zero program [43], based on a Swedish initiative, is aimed at eliminating pedestrian, cyclist, and driver deaths due to roadway accidents. Spatial data and analytics play a key role in the design of strategies to make streets safer. Aerial photography of street layouts along with traffic direction, turning rules, speed limits, past patterns of accidents and injuries, when combined and subjected to spatial analysis, allow transportation planners to identify dangerous intersections and street segments and to design strategies to mitigate risk. These can include turning restrictions, speed reductions, the use of bottlenecks and speed bumps, lane redesign, improved street lighting and signaling, traffic cams, dedicated bike lanes, and others. The city has recently reported that pedestrian fatalities have dropped 45% from 184 in 2013 to 101 in 2017 as a result of Vision Zero strategies [44].
- **Future Prospects:** Worsening urban traffic congestion that lengthens travel times and pollutes the environment, and the dangers inherent in mixing vehicles, pedestrians, and cyclists within the same crowded right-of-way, present transportation planners with their biggest challenges. One of the most promising ways of dealing with these problems is through the use of smarter vehicles. Transportation planners predict that through the use of spatially enabled sensor technologies, a significant portion of the job of moving through the city in a vehicle will be taken over by the vehicle itself, which will know where it is at every moment, while being continuously updated about the movements of other vehicles, objects, and people in its path. Real-time monitoring of traffic backups can provide drivers and vehicles with smarter routes to take and modify traffic signal timing to help unsnarl jams. In this new transportation environment based on real-time spatial detection, tracking, analysis, and automation, no vehicle will run a light or be blindsided or rear ended. Vehicle speed and direction will be automatically altered should a pedestrian or cyclist be at risk. While no one knows whether total automation of the driving function is possible, automated features are certain to make getting from one place to another more efficient and safer than it is now.

31.5.10 Environmental Planning

- **Background:** The health of our planet and the vitality of all its life sustaining natural systems is of vital concern to every individual. In particular, it is critical that we comprehensively understand the major environmental threats and design strategies that reduce threats. The collection, mapping, and analysis of spatially enabled information about our environment has a long history. Now, the use of a wide variety of location-aware sensors of great sensitivity and capacity, has revolutionized data collection and has enabled increasingly sophisticated and accurate analytics and modeling. At the same time, we are mapping the patterns of human development that impact the natural environment.
- **Smart Spatial Approaches:** We take the daily weather forecast for granted and many do not recognize that behind the generalized images of cold and warm fronts, storms, and sunny weather are millions of sensor readings and human observations that are run through spatial models to map current conditions and future predictions. Similarly, scientists measure the molecular composition of the atmosphere, water temperatures and ocean currents, the presence and effects of pollutants, and the geological characteristics of the earth's crust, including the movement of tectonic plates. The study of global warming is an excellent example of how spatial data about temperatures, the concentration of carbon dioxide, methane and other heat absorbing gases, and the volume of polar icecaps, among many other inputs, can be incorporated into complex models to help us understand rising temperatures,

changes in ocean levels, impacts on forests, and effects on agriculture. In addition to global monitoring and analytics, spatially enabled environmental systems are also being extensively deployed on the community level to deal with local water and air pollution problems.

- **Case Study:** NYC Watershed Management Program – The ability of New York City to exist at all, let alone to thrive, depends in large measure on its supply of fresh water. Evolving over the centuries since the city's founding, NYC's water supply now comes from a network of upstate reservoirs located within a watershed covering 2000 square miles. The system provides nine million people with fresh water – at a volume of 1.1 billion gal/day. The NYC Department of Environmental Protection (DEP) manages the City's water supply system. A top DEP priority is to keep the water clean and free of impurities and pathogens. Key to achieving this is a spatial monitoring and analysis system, managed by agency planners and scientists, that utilizes aerial photography, remote sensing, water quality and stream flow sensors, land use information, septic system mapping, and rigorous water quality testing. All this data is registered to an accurate basemap for analysis, enabling DEP to act swiftly to combat any threats to water quality. The success of this program has enabled the city to avoid the construction of a water filtration facility for its Catskill and Delaware systems estimated to cost in excess of $10 billion [45].
- **Future Prospects:** As increasing amounts of location enabled information becomes available about the natural systems that surround us and their interactions with the manmade systems that serve us, spatial analytics and modeling will be used for more resilient and sustainable designs. Optimal locations for solar and wind power will be found. Green roofs and smart sewers will lower the volume of stormwater runoff, reducing pollution while easing the burden on wastewater treatment plants. Access to electric vehicle refueling outlets will be made available at the curbside through the smart redesign of the electric grid. Efficient smart grid solutions for providing energy will supplement and eventually replace our current one-size-fits-all system, which is so heavily dependent on fossil fuels.

31.5.11 Smarter Businesses and Organizations

- **Background:** Spatially enabled smart cities start with the decision by county and municipal government leaders to build enterprise GIS systems as an extension to their IT infrastructure. The spatially enabled data generated by these municipal systems can then be placed in open data portals for access by businesses, nonprofit organizations, community groups, and citizens. In this way, Smart Cities extend the benefits of their spatial data assets to everyone, increasing overall levels of smartness, allowing organizations and individuals to find their own creative uses of these information assets. Additionally, some jurisdictions are teaching their students how to access and use open data as part of geography and civics curricula, encouraging them to devise projects that map vital characteristics of their communities and teaching them information skills that can be applied in their future work lives.
- **Smart Spatial Approaches:** Many private companies find that spatially enabled public and commercial data is indispensable for a variety of business functions, including locating facilities, marketing, supply chain tracking, product delivery, weather prediction, demographic analysis, real estate analysis, facility design and construction, and infrastructure management. The easier it is for businesses to find and use the information they need to improve productivity and profitability, the more likely they will flourish, expand, and provide additional jobs. Not-for-profit and community organizations have spatial information needs that are similar to those of private firms. Many are concerned about a particular neighborhood or focus on specific causes. They often require land use, demographic, infrastructure, and economic data to support their efforts. Open data and the availability of easy to use spatial tools can make them more effective and enhance the benefits they can contribute to their communities.
- **Case Study:** Vizalytics Inc. [46] is a private company that started in New York City and plans to expand to Chicago, Seattle, and San Francisco. It was founded in the aftermath of Hurricane Sandy in 2012, which affected tens of thousands of small business across the city. The Vizalytics team saw the need to make critical city-wide information easier to access and understand and developed the "Mind My Business" application. This app taps into a wide variety of city open datasets including information about construction, traffic, regulations, health and safety, fines, events, and 311 information. It helps shopkeepers and business owners make better decisions and identifies opportunities for them to save time and money. The application enables clients to tap into information directly pertaining to their location, using a friendly map-based user interface. In addition, Vizalytics, in partnership with the NYC Mayor's Office has built the Neighborhoods.nyc website [47], which provides information about more than 300 neighborhoods. The site is a resource for residents and a catalyst for community organizing and change.
- **Future Prospects:** We suggested earlier that there are four major sources of spatially enabled data that are increasing in type, volume, accuracy, and interoperability. Data sources include: enterprise spatial systems; remote, mobile and fixed sensors; personal data, including smartphones, wearables and social media; and the Internet of

Things (IoT). As this data becomes increasingly available and interoperable, and as governments and entrepreneurs learn to use this data for a variety of public and commercial purposes, the level of smartness in a jurisdiction will rise. There will be improved public services, more informed citizens, better educated students, and a more favorable climate for new businesses and more jobs.

31.5.12 Conclusion

In the ten use categories described above, we identified spatially enabled and spatially dominant smart city applications that are currently being used and those that are still on the horizon but are approaching rapidly. As increasing amounts of high-quality, standards-based, interoperable, spatially enabled information becomes available, and as improvements in analytics, models, and technologies are made, we see no end to the opportunities for improving our cities and our world through smarter and smarter spatial applications.

31.6 Return on Investment

31.6.1 Quantifying Benefits of Smart Spatial Systems

It is a challenge to quantify the benefits that are currently being achieved by smart spatial systems and to estimate the benefits expected to be realized in the future. Spatial technologies, while critically important to the success of many applications, rarely stand alone and must work together with other, nonspatial capabilities. This makes it difficult to isolate the benefits attributable to the spatial components.

Our approach will be to focus on the municipal and county government sector. This is because government work at the local level is dominated by geographically oriented service delivery to the public. Consequently, local governments have consistently been early adopters of spatial data infrastructures. Since there are great similarities in local government services across the US, we believe that the benefits documented by specific local examples have general applicability. We decided to examine three major categories of benefits: increased revenues, lives saved, and improved productivity. In the following, we will document a range of benefits and then develop default benefit values based on conservative estimates.

GISCalc [48]: An important source of benefits information comes from GISCalc, a cost-benefit analysis tool built by the Fund for the City of New York [49] for the New York State GIS Association to help local governments understand the value of spatial systems. GISCalc has brought together documentation of ROI (return on investment) from about 100 spatial applications organized by use categories.

31.6.2 Improved Revenue Collection:

When thinking about financial benefits achieved by local governments through the use of spatially enabled systems, ROI can be found *both* in enhanced revenue collections and in more productive municipal service delivery operations. We will start on the revenue side.

Property Taxation According to the Urban Institute [50] property taxes comprise 30% of the local government general revenue stream, which includes funding for county and municipal boards of education. We all know the importance to real estate of location, location, location. It is, therefore, not surprising that spatial capabilities have been aiding property tax assessment and collection work processes for decades (see Property Assessment and Taxation Use Case, Sect. 31.5.2). The combination of an accurate parcel and structure map, with accompanying attribute data, and overhead and oblique imagery, among other data types, gives assessors essential information to work with and has led to a more complete inventory of building floors, floor area, and other building improvements.

GISCalc suggests a default value of 2% (with a range of 0.5–5.0%) for improved revenue collection due to spatial support systems for property tax assessment and collections based upon reporting between 2005 and 2006 from Cuyahoga County, Ohio; Washtenaw County, Michigan; and Citrus County, Florida. The sources of improved revenue are the identification of unreported property improvements and the finding of untaxed parcels. Benefits were also realized by improvements in billing addresses. We believe that this default value can be increased to 4% given the advances in street-level imagery and the mapping of building interiors using lidar. Improvements are also being made in the ability to create 3-D models of air rights and underground easements.

Other revenue related benefits of spatial systems included in GISCalc include:

- **Charges [51] – water and sewer taxes.** Charges are a major category of municipal revenue and are composed of fines and fees including payments for water and sewer usage, parking violations, tolls, parking meters, real estate transactions, and a variety of other fees for financial and commercial activities. According to the Urban Institute, in 2015 charges accounted for 18% of local revenues and 10% of state revenues. GISCalc sets a default value of 2% for gains in water and sewer tax collections based on the identification of untaxed properties with service connections.
- **Charges – other tax billing and collection activities.** The effectiveness of billing and collection operations within a jurisdiction is affected by the completeness and accuracy of a central street name and address database

without which billing address error rates can be 10% or higher. Incorrect, inconsistent, and incomplete addressing information can make it difficult for municipal authorities to pursue those who are avoiding payment. However, a high-quality address database built into a geocoding application will help guarantee that address intake is accurate, bills get to their proper destination, and collection agents have good location data to work with. Additionally, by matching names and addresses across different revenue collection work processes, repeat scofflaws can be identified for targeted action. We believe that revenue increases of 1% can be realized for these operations.

- **Transfer payments – population-based block grants.** Transfer payments from federal and state governments [52] make up an average of 36% of total local government revenues. The Federal General Accounting Office reports that in fiscal year 2000, 85% of federal government obligations in grants to state and local governments were distributed on the basis of formulas that use data such as state population and personal income. GISCalc identifies a default gain of 3% in census counts due to the use of spatial data and analysis in undercounted areas. The default value of each additional headcount is $150 (although it could be considerably more than this). For a city with a population of 1 million, this amounts to $4,500,000.

Quantification We conclude that when effectively utilized, spatial systems play a major role in enhancing revenue collected by municipalities. At the present time, we estimate that municipalities that have built accurate property databases and are using available spatial tools are realizing a 2% gain in property tax revenue (which represents a 30% share in total revenue), amounting to an overall revenue gain of 0.6%. We believe that property tax revenue increases can double to 4% with the application of advanced spatial techniques including street-level and interior lidar, and 3-D spatial visualization and analytics, resulting in an overall increase of revenues of 1.2%. We also suggest adding an additional 0.4% for other revenue benefits that can be realized using currently available methods, and an estimated 0.8% gain with future improvements. This gives us overall revenue gains of 1% with current technology and an estimated 2% in the future. While these percentages appear small, when multiplied by total revenues collected, they amount to a very large amount of money.

31.6.3 Saving Lives

Spatial systems have a well-earned reputation for being essential components of applications that save lives. The saving of lives is often a function of the speed with which the location of a victim can be given to responders, the identification and dispatch of a nearby response team, and the time it takes for trained and well-equipped personnel to arrive at the scene to administer assistance.

Emergency response – ventricular fibrillation 911 emergency response systems are the primary way that citizens call for help when there is a dire need for assistance. Spatial capabilities are at the heart of 911 emergency response systems and are essential elements in reducing response times. In particular, it has been documented that lowering response times can reduce fatalities in cases where people suffer ventricular fibrillation. "Ventricular fibrillation (V-Fib) is the most serious cardiac rhythm disturbance. The lower chambers quiver and the heart can't pump any blood, causing cardiac arrest" [53]. GISCalc identifies a series of articles published in 2006 by USA Today and written by Robert Davis, who reported that there is one incident of V-Fib annually for every 5000 people. V-fib is almost always fatal if trained responders are not on the scene within 6 min [54].

Emergency response – additional categories of life saving Just as reduced 911 response times enable emergency responders to save the lives of victims of V-Fib, they also are known to reduce fatalities and limit the severity of injuries in the following additional types of incidents: gunshot wounds, fires, severe strokes and heart attacks, asthma and other types of breathing disorders, and vehicular accidents resulting in bleeding and other trauma. We know that in cases of opioid overdoses, when the drug naloxone is administered quickly, lives are saved. The New York Times in a November 25th, 2018 article notes that wider availability of naloxone could prevent 21,000 deaths over the next decade [55].

Based on the above, we feel safe in claiming that spatial systems, and their ability to precisely locate people at extreme risk, and to rapidly direct emergency personnel to an incident scene, have the potential to save one life in every 5000 city residents on an annual basis. For an urban area of one million people, this means saving 200 lives annually. Just for the sake of reference, GISCalc notes that a number of federal agencies including OMB, EPA, FDA, and DOT put the value of a human life between 6.0 million and $9.4 million [56].

Other ways that spatially enabled applications and analytics save lives

- **CompStat and intelligence led policing.** Police departments across the US use CompStat (computer statistics) programs to reduce major crime. In 1990, just before CompStat was introduced, 2245 New Yorkers were murdered. In 2017 the number of murders had fallen to 292, a reduction of 87%. Some of this decline can be credited to the use of CompStat enabled spatial analytics that examine patterns of violent incidents on a block-by-block basis and utilize that intelligence to fine tune police patrol tactics. While it is impossible to attribute a definitive percentage of reduced murders to the use of CompStat, intelligence led policing is recognized as an integral part of police operations and strategy.

- **Vision Zero.** The reader is referred to the Transportation Use Case (Sect. 31.5.8). In the case of Vision Zero, spatial analytics play a central role in identifying traffic hazards and designing mitigation strategies that have saved lives. Since Vision Zero was implemented in NYC, annual pedestrian fatalities have dropped from 184 in 2013 to 101 in 2017, a drop of 45%.
- **Epidemiology.** The reader is referred to the Health Use Case (Sect. 31.5.7). Spatial-based epidemiological studies and disease containment strategies have been utilized ever since Dr. John Snow created a map of a neighborhood in London, which tied an outbreak of cholera to contaminated water coming from a water pump. More recently, medical authorities have used advanced spatial analytics to control West Nile Virus outbreaks by identifying and treating hotspots where the disease is active. This has been shown to reduce human cases and save lives. The use of disease mapping and related spatial analytics is now a standard procedure for dealing with outbreaks of contagious diseases. It is used to track season flu outbreaks and several years ago it was key to helping bring the Ebola outbreak in West Africa under control. Spatial analytics is also used to assess the effects of environmental factors, like high levels of pollution, on disease and death.

How lives will be saved in the future We can envision a future where the use of wearables and smart phones to monitor individual health will allow the sudden onset of a life-threatening event or illness to be detected early, providing automatic notification to the wearer or triggering a call to first responders. We also anticipate smarter roadways and smarter vehicles that can facilitate the movement of emergency units (perhaps using smaller and more maneuverable ambulances) to reach victims in far less time than it now takes.

We conclude with the conviction that spatial systems geared to reducing fatalities are already saving the lives of at least 1 in 5000 citizens annually when emergency response times approach or drop below 6 min. This equates to the potential for saving 200 lives annually for every 1,000,000 people. We can easily see the rate of lives saved increasing to 1 in 2500 citizens (400 lives saved annually per 1,000,000 people) as new medical sensors and spatial technologies become available. Our calculations do not include the value of fewer and less severe injuries and reduced damage to property made possible when response times are lowered.

31.6.4 Improving Government Operations

Spatially enabled computer applications have an established record of improving government operations. Types of improvements, other than those related to increased revenues, include higher productivity, greater efficiency and effectiveness, cost reduction, and cost avoidance. We will examine three categories of municipal work and identify the spatially enabled applications that can lead to improvements. The three categories are field operations, administrative operations, and infrastructure management – a special area that combines both field and administrative work and is funded by both expense and capital budgets. It is our opinion that about 75% of municipal expenditures are used to support these three categories of work.

Field Operations

A significant part of municipal budgets are spent on operations that deliver services to customers where they live and work. Services include garbage collection, street cleaning, pothole repair, snow clearance, park and playground management, health and social services, air and noise complaint response, asset management, fire protection, public safety patrol and response, and health and food inspections. Spatially enabled systems provide special capabilities to support these and other kinds of field operations, including:

- Designing efficient routes for vehicles, equipment, and work teams lowering distances traveled and reducing fuel usage and vehicle wear and tear
- Enabling workers in the field to wirelessly access information related to their work location, reducing the need for return trips to the office.
- Allowing information to be accurately collected and processed, by location, in the field, reducing back-office administrative chores.
- Using detailed, up-to-date and even real-time remotely sensed data, allowing property assessments and other work processes to be reliably performed in the office instead of requiring field visits.
- Flagging duplicate complaints and bad locations to reduce wasteful trips.

Quantification GISCalc [57] includes a Field Efficiency Support category that contains 21 examples of how spatially enabled applications improve field operations. Nine of these use cases provided metrics in the form of percentage gains in productivity. Improvements included $\approx 1.5\%$ for reduced travel time for Cleveland park workers due to improved routing, and 37.5% for Fairfax County Virginia's use of GIS to optimize solid waste and recycling collection routes. Eight of the nine use cases, documented improvements of $> 6\%$.

Understanding that not every service delivery operation can realize benefits from spatial systems and wishing to be conservative in our estimates, we believe it is reasonable to claim that municipal governments can achieve an overall 3% improvement in field work productivity. As new and improved spatial systems become available, we expect this number to climb higher.

Back Office Operations Support, Planning, and Administration

A significant portion of the municipal workforce is engaged in field operations support, public interface, records management, planning and analysis, and supervisory functions. Municipal personnel have office space in government buildings and are equipped with computer and communications systems. Spatial systems streamline their work in a number of ways including:

- **311 and 911 systems.** 311 customer relationship management (CRM) and 911 emergency response systems have significantly changed the way municipal services are provided to the public. These systems allow callers to use just one three-digit phone number to get access to the services provided by dozens of government agencies. These systems use a comprehensive, standardized street name and address database to ensure location information is captured accurately and that the correct service yards, precincts, or fire districts are contacted, and the nearest field units are dispatched. Standardized location information ensures that data related to different services is interoperable and can be analyzed across service categories. Staff efficiencies are realized by centralizing call intake, freeing agency staff to concentrate on service delivery.
- **Address lookup.** Much of the back-office work conducted by municipal agencies, such as sending out tax bills, transferring property ownership, and approving building permits, involves recording accurate address information. This is another instance where having a standardized street name and address database can save enormous amounts of time while guaranteeing that location information is complete, accurate, and interoperable.
- **Field automation reduces workloads.** Because service workers have greater access to spatially enabled information in the field, they can fulfill paperwork requirements without the need to task clerical staff.
- **Citizen access to data.** Office personnel in many agencies once spent a great deal of time answering phone calls from the public. Now, web-enabled spatial applications allow citizens to look up information on their own, or to conduct transactions, often with the use of intuitive map interfaces.
- **Powerful planning tools.** Operation planners and analysts now have many more tools at their disposal. Routing algorithms, pattern analysis, and real-time monitoring of field teams and vehicles enable increases in work efficiency and improved supervision.

Quantification GISCalc [58] contains 15 use cases under the Office Efficiency Support category. Each describes improvements in office work due to spatially enabled systems, with four documenting the percentage improvement with an average of 28%. Because of the small sample size, we believe it is fair to estimate that the use of spatial enhanced applications results in an efficiency improvement of 3% or more.

Infrastructure Management

The capital budgets of municipal governments provide for long-term investment in facilities and infrastructure that can include school buildings, water, and sewer pipelines; roadways, tunnels and bridges; parks and cultural facilities, and buildings that house municipal operations. For a project to be "capital eligible" it must have a useable life of at least 5 years, but in many cases, lifetimes of 50 years or more are expected. Capital projects go through three major phases. Planning and design: when planners, architects, and engineers put together the details of what the project will look like and how it will function. Construction: the process of building the project. Operations and maintenance: the management of the facility during its lifetime, including facility upkeep and repair. Increasingly, municipalities are turning to asset management software packages to identify and locate each facility feature and track maintenance work. Proper maintenance can lengthen asset life and, over time, make capital programs more cost effective and efficient. Even a 2 or 3-year extension in facility life can result in significant capital savings and could result in lowered borrowing expenses.

The Field Efficiency Support category in GISCalc documents improvements that can be achieved during the operations and maintenance phase of capital construction projects. Municipalities and counties cited include: Cleveland, Ohio; Cuyahoga County, Ohio; Erie County, NYS; Miami-Dade County, Florida; and Honolulu, Hawaii. Average gains are over 3%.

Below we reference two, large-scale, federally funded research projects that document construction and maintenance-related cost savings from the perspective of improved data interoperability, quality, and completeness.

Data interoperability A study conducted by Michael P. Gallaher et al. *entitled Cost Analysis of Inadequate Data Interoperability in the US Capital Facilities Industry* for the National Institute of Standards and Technology (NIST), August 2004 [59], examined inefficiencies when data generated at different phases of construction and between different kinds of contractors was not shared. Failure to share was caused by lack of data standards and a lack of sharing procedures built into the construction process. We believe the NIST findings apply to all municipal construction projects whether in the form of above-ground buildings or underground infrastructure. The primary means of capturing interoperable building data is to apply the standards developed for building information modeling (BIM) that shows how all building components are networked and related to each other, an inherently spatial task. The best way to capture interoperable underground infrastructure data is to adopt common spatial data standards across utilities. In Flanders, Belgium, all utility companies are required to locate their underground assets in a standardized format making the data interoperable. In Auckland, New Zealand, there is a 3-D interoperable data model for its water and sewer networks that also connects with building information. Both Flanders and Auckland are realizing significant benefits, including reduced approval time for excavations permits and better excavation safety. The NIST study finds that improved data interoperability for construction projects yields a 1.8% savings across planning and design, construction, and maintenance phases.

Data accuracy and completeness Purdue University conducted a study for the Federal Highway Administration entitled *Cost Savings On Highway Projects Utilizing Subsurface Utility Engineering (SUE)* [60]. The study demonstrated that by increasing the accuracy of underground utility data from Quality Levels C and D to Quality Levels A and B, highway construction projects could save 1.9%. Similarly, improvements in utility data and interoperability would make excavations for routine maintenance, safer and faster, while avoiding accidental strikes. London, NYC, and Flanders, for example, are subjected to more than 200,000 annual street excavations. The Purdue study examined 71 projects with a combined construction value of $1 billion. This study presents a clear argument for accurately documenting the location of underground infrastructure in 3-D because failure to do so increases costs due to unexpected interferences, dangerous utility strikes, and construction delays. We believe that the results of this study can be extended to all building and facility construction projects.

Quantification Through the use of asset management systems, and improvements in data accuracy, completeness, and interoperability (using SUE, BIM), we believe that lifecycle infrastructure construction and maintenance costs can be reduced by a conservative 3%. We believe that for municipal capital design, construction and management programs, costs can be reduced by a default value 3%, knowing that in many instances they will be higher. We also believe that as new and improved infrastructure sensing and monitoring technologies are developed, and as more complete and interoperable data models are designed and deployed, the potential for improvements in the capital sector can rise to 5%, even without taking into consideration the potential benefits of extended facility life.

Additional considerations The studies cited below provide additional support for cost savings in this area.

Imagining Construction's Digital Future by Rajat Agarwal, Shankar Chandrasekaran, and Mukund Sridhar, McKinsey and Company [61]: "The construction industry is ripe for disruption. Large projects across asset classes typically take 20% longer to finish than scheduled and are up to 80% over budget. Geological surprises are a major reason that projects are delayed and go over budget. Discrepancies between ground conditions and early survey estimates can require costly last-minute changes to project scope and design. New techniques that integrate high-definition photography, 3-D laser scanning, and geographic information systems, enabled by recent improvements in drone and unmanned aerial vehicle (UAV) technology, can dramatically improve accuracy and speed."

Natural Hazard Mitigation Saves, 2017 Interim Report by the National Institute of Building Sciences (NIBS) [62]: Disasters and large-scale emergencies and accidents can have a large impact on municipal infrastructure and the built environment. The National Institute of Building Sciences (NIBS) reports that there is a $6 savings for every $1 invested in disaster mitigation projects. We believe that among the most effective disaster mitigation projects is building a base of integrated spatial information that can be used for analyzing and modeling infrastructure threats, identifying single points of failure and triggers for cascading effects, and implementing strategies to minimize risk.

31.6.5 Summary of Benefit Findings

The chart below aggregates the benefits we estimate are possible if spatially enabled systems are implemented to their full capacity (Table 31.1). The column of numbers beneath the *Currently feasible annual benefit rate* heading are benefit rates achievable if current state of the art methods and technologies are utilized. The column of numbers beneath the *Future potential benefit rate* heading are benefit rates likely to be achievable over the next 5 or 10 years, as new, improved, and smarter spatial methods and technologies are put to use. Readers should feel free to apply these benefit estimates to the known expense and capital budgets of local governments of interest to them.

The kinds of investment required of local governments for enterprise spatial systems as described in Sect. 31.4, while significant, build upon and extend existing IT infrastructure and create the foundation for hundreds of government applications, and for use by private companies, NGOs, and

Table 31.1 Summary of estimated current and future benefits from spatial systems

Benefit category	Benefit quantification standard	Currently feasible annual benefit rate	Benefit estimate	Future potential annual benefit rate	Future benefit annual estimate
Increased total revenues	Per $1 billion currently collected	1%	$10 million	2%	$20 million
Lives saved	Lives saved per 1 million people	1/5000	200 lives saved	1/2500	400 lives saved
Value of operational improvements	Per $1 billion currently spent: expense and capital	3%	$30 million	5%	$50 million

Note: Benefit rates will vary by jurisdiction and will tend to reflect the quality of a jurisdiction's spatial data infrastructure. It is likely that the most spatially enabled cities can realize benefits that significantly exceed the default numbers we are suggesting.

community groups. We believe we have more than proven that the benefits stream produced by such investments will be repaid many times over. Should a municipality fail to move forward with an adequate spatial program, its ability to capitalize on Smart City programs and innovations will be seriously jeopardized.

31.7 Conclusion: Smart Cities Must Be Spatially Enabled

In this chapter, we have shown that spatial information systems and spatial capabilities have evolved out of a union of information technology and the science of geography and that they empower Smart City applications. We have identified the unique capabilities that allow spatially enabled systems to add significantly to the power of IT; in our opinion, as much as doubling the computing capabilities of nonspatial systems. We have identified the new types of high-volume spatially enabled data becoming available and noted that efforts are underway to create data models that promote data interoperability and analytics in areas like underground infrastructure, building information modeling, and disaster preparedness and response. We have emphasized the importance of spatial data standards and how spatial operations must be supported by an appropriate technology architecture with appropriate levels of resources and staffing. We then described current applications and future directions for a number of use categories. We make the point that for a large number of government agencies, spatially enabled applications are the "dominant" expression of IT. Finally, we quantified the benefits of spatially enabled applications, showing that they increase revenues, save lives and raise productivity. We believe that we have demonstrated that the success of a smart city program is inconceivable without the significant use of spatially enabled capabilities and applications.

Future Smart Cities In conclusion, we refer the reader to the study *Smart Cities and Cost Savings* by ABI Research, Chordant and CA Technologies, which estimates greater than $5.4 trillion annually in worldwide Smart Cities' savings, much of this amount to be realized by cities [63]. When set against the 2017 figure for world domestic product of $80 trillion [64], we can see that the implementation of Smart Cities programs and applications are expected to produce future cost savings of $\approx$ 6.75%. To the extent that this estimate is correct, we would add that the quality of a city's spatial data infrastructure will to a significant degree determine how much a share of these benefits any one city will realize.

Digital Twins for Smart Cities Digital Twins – a recent concept advancing integrated modeling – enables Smart Cities to integrate physical, digital and human activities in the built environment for the benefits of its citizens. A Digital Twin is a virtual representation of the real world including physical objects, processes, relationships, and behaviors [65]. Digital Twins are characterized by the coalescing and integration of a number of technologies that have rapidly evolved, including new generations of sensors, the Internet of Things (IoT), earth observation (EO), crowd sourcing, predictive modeling, artificial intelligence, and machine learning. Building on early adopters, municipal governments are extending their GIS infrastructure using Digital Twin technologies.

Application of Digital Twins at urban-scale builds on existing GIS capabilities. New data sources for building Urban Digital Twins includes advances in rapid mapping technologies, remote sensing, and laser scanning. GIS integrates these data sources to create 3D models that can be used to facilitate communication, understanding, and analysis of complex systems, such as entire cities. Successful Urban Digital Twin implementations depend upon a solid enterprise GIS foundation working with data and systems from the Architecture, Engineering and Construction (AEC) industry. BuildingSMART international and OGC are fostering collaboration [66] between the geospatial and the AEC domains to facilitate integration and interoperability of built environment data.

Looking further ahead, Digital Twins are combining predictive modeling with real-world monitoring. Internet of Things and other sensor-based information systems increase the availability of real-time data about devices, location, weather, traffic, people movement, etc. Combining this data with predictive models allows estimates of the status and prediction of the future state of urban systems. Digital Twin dynamic predictive modeling capabilities allows for resource management covering mobility, energy, and public safety. The COVID-19 pandemic raised sensitivity for and awareness of 24/7 operations of city services, such as water supply and sanitation; a task that Digital Twin technologies are predestined to support. Digital Twin solutions not only enable simulation and monitoring but are used in real-time by emerging applications to position themselves with sub-meter accuracy both indoors and out in order to solve critical management and decision support challenges. Successful dynamic Digital Twin implementations depend upon a solid GIS foundation that gives accurate spatial identity to all information.

Benefits of dynamic Urban Digital Twins are already seen in resource management for several cities. Improved energy management is crucial to adapting to climate change. Several cities have deployed energy Digital Twins that enable planning and operational management of energy consumption in their cities. The Helsinki Energy and Climate Atlas [67] contains a wide variety of information related to the energy, including solar energy potential, heating demand prediction, geo-energy potential, and energy data of buildings across the city. Energy mapping and urban scale building energy modeling makes it possible to analyze the status quo and to project scenarios for reducing carbon emissions. Using urban energy system modeling to reduce the causes of climate change will benefit next generation cities and the globe.

To achieve the desirable objectives of Digital Twins, institutional efforts for model construction and management are needed for use by municipal governments and their suppliers are being developed (e.g. [68]). Successful Digital Twin projects must navigate scope complexity, participant diversity, data ownership, and security. Using a system-of-systems approach, the variety of Digital Twins will be made interoperable, operational and maintainable. Ultimately, Digital Twins must help the owner or operator of the real-world asset or system to solve business or mission-critical applications to justify the investment and process changes that will inform and maintain the twin. Similar to the Spatial Data Infrastructures based on GIS, government leadership can create the conditions that will enable the technology and operations to develop and thrive for smarter cities. Moreover, given the huge advances in spatial analytics, this information can be used to predict the future including infrastructure aging and failures, the spread of disease, the impacts of natural disasters, and international challenges like global warming. The application of Digital Twins will significantly improve productivity and allow society to mitigate existential threats. The integrative and interoperable powers of spatially enabled data and technology, as expressed through Digital Twins, will be at the very center of these advances.

Stephen Levin, New York City Council Member, commented "as we enter the Digital Twin era, as befits an elected official, I am chiefly interested in how this technology will benefit people, making life safer, healthier and more prosperous. A number of applications will ensure basic building services, monitor air quality, reduce vehicle accidents, diagnose health problems, and support infrastructure planning and management. It is important to help non-technical leaders to understand Digital Twins so they can be better advocates for this extraordinary technology" [69].

Acknowledgements The writing of this chapter was made possible by the generous support of the Fund for the City of New York and its President, Dr. Mary McCormick; and by the Members of the Open Geospatial Consortium.

The authors would like to acknowledge the following individuals for their assistance in the writing of this chapter: Wendy Dorf, President NYGeoCATS; Jim McConnell, NYC Emergency Management; Jack Eichenbaum, Queens Borough Historian; Douglas Williamson, GIS Manager, NYPD.

References

1. BSI: PAS-180. (2014). https://www.bsigroup.com/en-GB/smart-cities/Smart-Cities-Standards-and-Publication/PAS-180-smart-cities-terminology/
2. 100 Resilient Cities: Safer and Stronger Cities. http://www.rebuildbydesign.org/data/files/918.pdf. pg.13
3. Wikipedia: Smart City. https://en.wikipedia.org/wiki/Smart_city
4. Barlow, M., Levy-Bencheton, C.: Smart Cities, Smart Future. John Wiley & Sons, Hoboken (2019). pg. xvii
5. World Bank: Urban Population. https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS
6. Milner, G.: Pinpoint. W.W. Norton & Company, New York (2016)
7. The Spatial Data Infrastructure Cookbook, Version 2.0 (2005). ftp://www.zcpwz.pl/Materialy_INSPIRE_12.13.0.2009/GSDI%20CookBook/cookbookV2.0.pdf
8. Harvey, F., Tulloch, D.: Local Government Data Sharing. https://www.tandfonline.com/doi/abs/10.1080/13658810600661607
9. Carrera, F., Ferreira Jr., J.: The future of spatial data infrastructures. Int. J. Spat. Data Infrastruct. Res. **2**, 49–68 (2007)
10. The Open Geospatial Consortium: The OGC's Role in Government and Spatial Data Infrastructure. www.opengeospatial.org/domain/gov_and_sdi
11. Wikipedia: Linked Data. http://en.wikipedia.org/wiki/Linked_data
12. Wikipedia: Semantic Query. https://en.wikipedia.org/wiki/Semantic_query
13. W3C Semantic Web: Resource Description Framework (RDF). https://www.w3.org/RDF
14. Wikipedia: Web Ontology Language (OWL). https://en.wikipedia.org/wiki/Web_Ontology_Language
15. Wikipedia: GeoSPARQL. https://en.wikipedia.org/wiki/GeoSPARQL
16. The Open Geospatial Consortium: Indexed 3D Scene Layers (I3S). https://www.opengeospatial.org/standards/i3s
17. The Open Geospatial Consortium: Discrete Global Grid Systems SWG. http://www.opengeospatial.org/projects/groups/dggsswg
18. insideBIGDATA: The Exponential Growth of Data. https://insidebigdata.com/2017/02/16/the-exponential-growth-of-data/

19. OGC, ISO, TC 211, IHO: Guide to the Role of Standards in Geospatial Information Management. https://ggim.un.org/meetings/GGIM-committee/8th-Session/documents/Standards-by-Tier-2018.pdf. p. 15
20. The Open Geospatial Consortium: OGC Standards and Supporting Documents. https://www.ogc.org/docs/is
21. The Open Geospatial Consortium: About OGC. https://www.ogc.org/about
22. The Open Geospatial Consortium: OGC Approves KML as Open Standard. http://www.opengeospatial.org/pressroom/pressreleases/857
23. The Open Geospatial Consortium: OGC Standards. http://www.opengeospatial.org/docs/is
24. The Open Geospatial Consortium: OGC SensorThings API. https://www.opengeospatial.org/standards/sensorthings
25. The Open Geospatial Consortium: CityGML. https://www.opengeospatial.org/standards/citygml
26. The Open Geospatial Consortium: Building Information Models (BIM) Demonstrated in OGC Testbed. http://www.opengeospatial.org/pressroom/pressreleases/732
27. The Open Geospatial Consortium: Underground Infrastructure Concept Development Study. http://www.opengeospatial.org/projects/initiatives/undergroundcds
28. Percivall, G. (ed.): OGC Smart Cities Spatial Information Framework. OGC White Paper, OGC document 14-115 (2015). https://portal.opengeospatial.org/files/?artifact_id=61188
29. The Open Geospatial Consortium: Smart City Interoperability Reference Architecture (SCIRA). http://www.opengeospatial.org/projects/initiatives/scira
30. New York City Open Data: NYPD Complaint Map. https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Map-Year-to-Date-/2fra-mtpn
31. Urban Institute: Property Taxes. https://www.urban.org/policy-centers/cross-center-initiatives/state-local-finance-initiative/projects/state-and-local-backgrounders/property-taxes
32. Computer Assisted Mass Appraisal. http://enacademic.com/dic.nsf/enwiki/11569894
33. New York State GIS Association: Emerging GIS, GIS Return on Investment Calculator, Property Tax Tab. https://www.nysgis.net/featured/emerging-gis-resources/
34. Digital Trends: Google Teams Up With 911 To Locate Emergency Callers More Easily. https://www.digitaltrends.com/mobile/google-911-accurate-location-data/
35. National Emergency Number Association (NENA): NG9-1-1 Project. https://www.digitaltrends.com/mobile/google-911-accurate-location-data/
36. Wikipedia: CompStat. https://en.wikipedia.org/wiki/CompStat
37. Wikipedia: Real Time Crime Center. https://en.wikipedia.org/wiki/Real_Time_Crime_Center
38. N. Singer: Mission Control, Built for Cities. The New York Times, 3 Mar 2012. https://www.nytimes.com/2012/03/04/business/ibm-takes-smarter-cities-concept-to-rio-de-janeiro.html
39. Daems, J.: KLIP: A radical, INSPIRE based digital exchange platform for utility network information in Flanders (2016). https://inspire.ec.europa.eu/events/conferences/inspire_2016/schedule/submissions/237.html
40. Wikipedia: Dr. John Snow. https://en.wikipedia.org/wiki/John_Snow
41. Center for Advanced Research of Spatial Information, Hunter College. http://carsi.hunter.cuny.edu/
42. Theophilides, C.N., Ahearn, S.C., Binkowski, E.S., Paul, W.S., Gibbs, K.: First evidence of West Nile virus amplification and relationship to human infections. Int. J. Geogr. Inf. Sci. **20**(1), 103–115 (2006)
43. New York City Government: Vision Zero. https://www1.nyc.gov/site/visionzero/index.page
44. New York City Government: De Blasio Administration Releases Annual Vision Zero Report (2018). https://www1.nyc.gov/office-of-the-mayor/news/156-18/de-blasio-administration-releases-annual-vision-zero-report
45. New York City Government: Watershed Protection. http://www.nyc.gov/html/dep/html/watershed_protection/index.shtml
46. Vizalytics. https://www.vizalytics.com/
47. Neighborhoods.nyc. http://www.neighborhoods.nyc/welcome.html
48. New York State GIS Association: Emerging GIS, GIS Return on Investment Calculator, Property Tax Tab. https://www.nysgis.net/featured/emerging-gis-resources/. Developers: John Wysel, Paul Epstein, Alan Leidner
49. Fund for the City of New York. http://www.fcny.org/fcny/
50. Urban Institute: State and Local Revenues – Local General Revenue. https://www.urban.org/policy-centers/cross-center-initiatives/state-local-finance-initiative/state-and-local-backgrounders/state-and-local-revenues
51. Urban Institute: State and Local Revenues – Charges. https://www.urban.org/policy-centers/cross-center-initiatives/state-local-finance-initiative/state-and-local-backgrounders/charges
52. Urban Institute: State and Local Revenues – Transfers. https://www.urban.org/policy-centers/cross-center-initiatives/state-local-finance-initiative/state-and-local-backgrounders/state-and-local-revenues
53. American Heart Association: Ventricular Fibrillation. http://www.heart.org/en/health-topics/arrhythmia/about-arrhythmia/ventricular-fibrillation
54. R. Davis: Six minutes to live or die. USA Today, 20 May 2005. https://usatoday30.usatoday.com/news/nation/ems-day2-cover.htm
55. A. Goodnough: The treatment gap. The New York Times, 25 Nov 2018. https://www.nytimes.com/2018/11/25/health/opioid-overdose-deaths-dayton.html
56. F. Partnoy: The cost of a human life – Statistically speaking. The Globalist, 21 July 2012. https://www.theglobalist.com/the-cost-of-a-human-life-statistically-speaking/
57. New York State GIS Association: Emerging GIS, GIS Return on Investment Calculator, Field Efficiency Support Tab. https://www.nysgis.net/featured/emerging-gis-resources/
58. New York State GIS Association: Emerging GIS, GIS Return on Investment Calculator, Office Efficiency Support Tab. https://www.nysgis.net/featured/emerging-gis-resources/
59. Gallaher, M.P., O'Connor, A.C., Dettbarn Jr., J.L., Gilday, L.T.: Cost Analysis of Inadequate Data Interoperability in the U.S. Capital Facilities Industry. NIST, Gaithersburg (2004). http://nvlpubs.nist.gov/nistpubs/gcr/2004/NIST.GCR.04-867.pdf
60. Purdue University: Cost Savings On Highway Projects Utilizing Subsurface Utility Engineering (SUE). FHWA, Washington (1999). https://www.fhwa.dot.gov/programadmin/pus.cfm
61. Agarwal, R., Chandrasekaran, S., Sridhar, M.: Imagining Construction's Digital Future (2016). https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/imagining-constructions-digital-future
62. National Institute of Building Sciences: Natural Hazard Mitigation Saves, 2017 Interim Report. NIBS, Washington (2017). https://www.nibs.org/news/381874/National-Institute-of-Building-Sciences-Issues-New-Report-on-the-Value-of-Mitigation.ht
63. ABI Research, Chordant and CA Technologies: Smart Cities and Cost Savings (2017). https://www.chordant.io/white_papers/abi-research-smart-cities-and-cost-savings
64. Statista: Global Gross Domestic Product. https://www.statista.com/statistics/268750/global-gross-domestic-product-gdp/

65. Digital Twin Consortium. https://www.digitaltwinconsortium.org/hot-topics/the-definition-of-a-digital-twin.htm (accessed: May 12, 2021)
66. "Built environment data standards and their integration," OGC and BuildingSmart, 02 March 2020, OGC Document 19-091r1, bSI document TR1012. https://portal.ogc.org/files/92634
67. City of Helsinki, Energy and Climate Atlas. [Online]. Available: https://kartta.hel.fi/3d/atlas/#/ (accessed: May 12, 2021)
68. The National Digital Twin programme, Centre for Digital Built Britain. https://www.cdbb.cam.ac.uk/what-we-do/national-digital-twin-programme (accessed: May 12, 2021)
69. Stephen Levin, keynote address, Location Powers: Urban Digital Twins summit, January 12-15, 2021. https://www.locationpowers.net/events/2101urbanvirtual/ (accessed: May 12, 2021)

Alan Leidner Alan Leidner is the Director of the Center for Geospatial Innovation at the Fund for the City of New York. In that position, he supports the underground infrastructure data interoperability project of the Open Geospatial Consortium. He has a degree in Urban Planning from Brooklyn's Pratt Institute and worked for 35 years as a planner and manager with New York City government.

George Percivall George Percivall is the Principal of GeoRoundtable which provides geospatial information engineering consulting. Previously he was CTO and Chief Engineer of the Open Geospatial Consortium; Chief Engineer with Hughes Aircraft for NASA's Earth Observing System Data and Information System – Landsat/Terra release; and Principal Engineer for NASA's Digital Earth Office and represented NASA in OGC, ISO, and CEOS. He holds a BS in Engineering Physics and an MS in Electrical Engineering from the University of Illinois-Urbana.

Glossary

a posteriori classification classification scheme based upon definition of classes after clustering the field samples collected
[ISO 19144-1:2009]

a priori classification classification scheme structured so that the classes are abstract conceptualizations of the types actually occurring
NOTE 1 The approach is based upon definition of classes before any data collection actually takes place
[ISO 19144-1:2009]

abbreviation designation formed by omitting words or letters from a longer form and designating the same concept
[ISO 1087-1:2000, ISO 19104:2016]

absolute accuracy closeness of reported coordinate values to values accepted as or being true
NOTE 1 Absolute accuracy is stated with respect to a defined datum or reference system.
NOTE 2 Absolute accuracy is also termed "external accuracy".
[ISO 19159-2:2016]

abstract root <programming> common root classifier of a category which is a superclass of any other classifier in the category
NOTE 1 The class Any in some programming languages is the abstract root of all classes. Thus, it is the de facto union of all classes. In this document, Geometry is the (named and explicit) abstract root for all geometry objects. In the package Geometry and any of its subpackages (including those in its Requirements Classes), any interface will be a subtype of Geometry either directly or transitively.
[ISO 19107:2019]

abstract test case generalized test for a particular requirement
NOTE 1 An abstract test case is a formal basis for deriving executable test cases One or more test purposes are encapsulated in the abstract test case. An abstract test case is independent of both the implementation and the values. It should be complete in the sense that it is sufficient to enable a test verdict to be assigned unambiguously to each potentially observable test outcome (i.e. sequence of test events).
[ISO 19105:2000]

abstract test method method for testing implementation independent of any particular test procedure
[ISO 19105:2000]

abstract test module set of related abstract test cases
NOTE 1 Abstract test modules may be nested in a hierarchical way.
[ISO 19105:2000]

abstract test suite ATS
abstract test module specifying all the requirements to be satisfied for conformance
NOTE 1 Abstract test suites are described in a conformance clause.
[ISO 19105:2000]

acceptance testing <user> process of determining whether an implementation satisfies acceptance criteria and enables the user to determine whether to accept the implementation
NOTE 1 This includes the planning and execution of several kinds of tests (e.g. functional, volume, performance tests) that demonstrate that the implementation satisfies the user requirements.
NOTE 2 This is not a part of conformance testing
[ISO 19105:2000]

access point location where travellers can enter or exit a transfer node
NOTE 1 An access point may not be a stop point. An access point may for example be the entrance to a railway station or the connection between a parking area and a railway station.
[ISO 19147:2015]

access rights information information that identifies the access restrictions pertaining to the content information, including the legal framework, licensing terms, and access control
NOTE 1 Access rights information contains the access and distribution conditions stated within the submission agreement, related to both preservation (by the OAIS) and final usage (by the consumer). It also includes the specifications for the application of rights enforcement measures.
[ISO 14721:2012, ISO 19165-1:2018]

W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*, Springer Handbooks,
https://doi.org/10.1007/978-3-030-53125-6

access software type of software that presents part of or all of the information content of an information object in forms understandable to humans or systems
[ISO 14721:2012, ISO 19165-1:2018]

accessibility ability to access and benefit from the functionality provided by a service or a facility
NOTE 1 Accessibility is often associated with disabilities. According to the concept of universal design, accessibility is, however, a matter that permanently or temporarily is relevant to all of us, e.g. people with heavy luggage, people with broken legs, people with small children, elderly people, etc.
[ISO 19147:2015]

accessibility information information about accessibility issues
NOTE 1 According to the concept of universal design, accessibility information should be addressed in a neutral way, i.e. not directed towards people with specific disabilities.
[ISO 19147:2015]

accuracy of measurement measurement accuracy, accuracy closeness of agreement between a test result or measurement result and the true value
NOTE 1 The concept "measurement accuracy" is not a quantity and is not given a numerical quantity value. A measurement is said to be more accurate when it offers a smaller measurement error.
NOTE 2 The term "measurement accuracy" should not be used for measurement trueness and the term measurement precision should not be used for "measurement accuracy", which, however, is related to both these concepts.
NOTE 3 "Measurement accuracy" is sometimes understood as closeness of agreement between measured quantity values that are being attributed to the measurand.
[ISO 6709:2008]

active object object which is capable of independent actions, and therefore capable of initiating interactions between itself and other objects without immediate prior external stimulation
NOTE 1 cf. passive object. An active object can represent a user, or an active service that depends on internal (and therefore not visible) triggers to start actions. Active and passive states can exist for the same object, and such a service can transition between these two states depending on invocation of an activation or deactivation operation protocol.
[ISO 19132:2007]

active sensing system sensing system that emits energy that the sensor uses to perform sensing
[ISO 19130-1:2018]

active sensor sensor that generates the energy that it uses to perform the sensing
[ISO 19130-2:2014]

active sonar type of active sensor that transmits sound waves into the water and receives the returned waves echoed from objects in the water
[ISO 19130-2:2014]

address structured information that allows the unambiguous determination of an object for purposes of identification and location
EXAMPLE 1 Address where the object is a business: 611 Fifth Avenue, New York NY 10022.
EXAMPLE 2 Address where the object is a building: Lombardy House, 809 Lombardy Street, The Hills, 0039, South Africa.
EXAMPLE 3 Address where the object is a land parcel for a building: San 4–5, Munjae-ro, Songpa-gu, Seoul, 13144, South Korea.
EXAMPLE 4 Address where the object is a building group, such as a school or large apartment area: 228-dong 404-ho, 26 Kyunghee-daero, Dongdaemun-gu, Seoul 130–701, South Korea.
NOTE 1 The object is identifiable in the real world, i.e. electronic and virtual addresses are excluded.
NOTE 2 "Identification" refers to the fact that the structured information in the address unambiguously determines the object, i.e. it helps the human to identify the object. In other words, "identification" here does not refer to unique identifiers in a database or dataset.
NOTE 3 There can be many addresses for an object, but at any moment (or lifecycle stage), an address unambiguously determines a single object.
NOTE 4 Two addresses from two different address classes (i.e. they have different sets of components) for the same addressable object are two different addresses.
NOTE 5 Two addresses for the same addressable object and from the same address class, but in two different languages are two different addresses.
NOTE 6 In addition to the addressable object, there may be a multitude of people, organizations, addressees or other objects associated with an address. These are external to the address model.
[ISO 19160-1:2015]

address alias one of a set of addresses unambiguously determining the same addressable object
[ISO 19160-1:2015]

address class description of a set of addresses that share the same address components, operations, methods, relationships, and semantics
EXAMPLE 1 "25 Blue Avenue Hatfield 0028" and "384 Green Street Motherville 2093" are from the same address class.

EXAMPLE 2 "PO Box 765 Goodwood 33948" and "PO Box 567 Grayville 98373" are from the same address class.
[ISO 19160-1:2015]

address component constituent part of the address
NOTE 1 An address component may reference another object such as a spatial object (e.g. an administrative boundary or a land parcel) or a non-spatial object (e.g. an organization or a person).
NOTE 2 An address component may have one or more alternative values, e.g. alternatives in different languages or abbreviated alternatives.
[ISO 19160-1:2015]

address position position representing the address
NOTE 1 An address may be represented by more than one position, e.g. different entrances to a building.
[ISO 19160-1:2015]

address reference system defined set of address components and the rules for their combination into addresses
[ISO 19160-1:2015]

addressable object object that may be assigned an address
[ISO 19160-1:2015]

addressee party who is the ultimate recipient of a delivery item or service
NOTE 1 The addressee may be explicitly defined as part of the postal address, or may be implicit. For example, in certain countries, omission of addressee information is taken as implying that delivery is to be to an individual or legal entity having legal access to the delivery point.
NOTE 2 Mr. or Mrs. Smith specifies that the addressee is either of two individuals, while Mr. Jones and Mrs. Smith denote that the addressee is a group of two individuals. See also role descriptor.
NOTE 3 The use made by the postal operator of addressee and mailee data might be dependent on the postal service applicable to the postal item. For some services, such as registered mail, the postal operator's responsibility might include ensuring that the addressee or a duly authorized representative acknowledges receipt of the postal item. In other cases, addressee data could be purely informative or used by the postal operator only for consistency checking and/or for the activation of forwarding services. In other cases, it might be used for sorting or sequencing purposes prior to delivery, e.g. in the case of business mail being pre-sequenced by department or individual company official.
NOTE 4 In some countries, the addressee may be an abstraction such as "postal customer".
[ISO 19160-4:2017]

addressing activities involving addresses
[ISO 19160-1:2015]

adjustable model parameters model parameters that can be refined using available additional information, such as ground control points, to improve or enhance modelling corrections
[ISO 19130-1:2018]

administrative source source with the administrative description (where applicable) of the parties involved, the rights, restrictions and responsibilities created and the basic administrative units affected
EXAMPLE 1 It is the evidence of a party's right to a basic administrative unit.
EXAMPLE 2 A document describing a transaction (a deed), or a judgement of the register holder.
[ISO 19152:2012]

admitted term term rated according to the scale of the term acceptability rating as a synonym for a preferred term
[ISO 1087-1:2000, ISO 19104:2016]

affine coordinate system coordinate system in Euclidean space with straight axes that are not necessarily mutually perpendicular
[ISO 19111:2019]

aggregation <UML> special form of association that specifies a whole-part relationship between the aggregate (whole) and a component part
NOTE 1 See composition
[UML 1, ISO 19103:2015]

AIP edition AIP whose content information or preservation description information has been upgraded or improved with the intent not to preserve information, but to increase or improve it
NOTE 1 This definition only refers to digital migration. NOTE 2 An AIP edition is not considered to be the result of a migration.
[ISO 14721:2012, ISO 19165-1:2018]

AIP version AIP whose content information or preservation description information has undergone a transformation on a source AIP and is a candidate to replace the source AIP
NOTE 1 An AIP version is considered to be the result of a digital migration.
[ISO 14721:2012, ISO 19165-1:2018]

along-track direction in which the sensor platform moves
[ISO 19130-1:2018]

altitude height where the chosen reference surface is mean sea level
[ISO 6709:2008]

ambient intelligence convergence of ubiquitous computing, ubiquitous communication, and interfaces adapting to the user
[ISO 19154:2014]

annotation any marking on illustrative material for the purpose of clarification

NOTE 1 Numbers, letters, symbols, and signs are examples of annotation
[ISO 19117:2012]

annotation <OWL> additional information associated to ontologies, entities, and axioms
[OWL, ISO 19150-2:2015]

annotation property <OWL> element used to provide a textual annotation for an ontology, axiom, or an IRI
[OWL, ISO 19150-2:2015]

antenna pattern ratio of the electronic-field strength radiated in the direction θ to that radiated in the beam-maximum direction
[ISO 19159-3:2018]

aperture reference point ARP
3D location of the centre of the synthetic aperture
NOTE 1 It is usually expressed in ECEF coordinates in metres.
[ISO 19130-1:2018]

application manipulation and processing of data in support of user requirements
[ISO 19101-1:2014]

application ontology ontology representing the concepts and relationships in an application schema
[ISO 19150-2:2015]

application schema conceptual schema for data required by one or more applications
[ISO 19101-1:2014]

arc <geometry> segment of a curve
[ISO 19107:2019]

archival information package AIP
information package, consisting of the content information and the associated preservation description information (PDI), which is preserved within an OAIS
[ISO 14721:2012]

area recording instantaneously recording an image in a single frame
[ISO 19130-2:2014]

association <UML> semantic relationship that can occur between typed instances
NOTE 1 A binary association is an association among exactly two classifiers (including the possibility of an association from a classifier to itself).
[UML 2, ISO 19103:2015]

associative concept system concept system based on associative relations
[ISO 19146:2018]

associative relation pragmatic relation
relation between two concepts having a non-hierarchical thematic connection by virtue of experience
NOTE 1 An associative relation exists between the concepts 'education' and 'teaching', 'baking' and 'oven'.
[ISO 1087-1:2000, ISO 19146:2018]

attitude orientation of a body, described by the angles between the axes of that body's coordinate system and the axes of an external coordinate system
NOTE 1 In positioning services, this is usually the orientation of the user's platform, such as an aircraft, boat, or automobile.
[ISO 19116:2019]

attribute named property of an entity
NOTE 1 Describes a geometrical, topological, thematic, or other characteristic of an entity
[ISO/IEC 2382:2015, ISO 19130-2:2014]

attribute <UML> feature within a classifier that describes a range of values that instances of the classifier may hold
NOTE 1 An attribute is semantically equivalent to a composition association; however, the intent and usage is normally different.
NOTE 2 "Feature" used in this definition is the UML meaning of the term.
[UML 1, ISO 19103:2015]

attribute <XML> name-value pair contained in an element
NOTE 1 In this document an attribute is an XML attribute unless otherwise specified. The syntax of an XML attribute is "Attribute::= Name = AttValue". An attribute typically acts as an XML element modifier (e.g. <Road gml:id = "r1" />; here gml:id is an attribute).
[ISO 19136-1:2020]

attribute event value of an attribute of a feature that may apply to only part of the feature
NOTE 1 An attribute event includes the linearly referenced location where the attribute value applies along the attributed feature
NOTE 2 An attribute event may be qualified by the instant in which, or period during which, the attribute value applied.
[ISO 19148:2012]

attributed feature feature along which an attribute event applies
[ISO 19148:2012]

azimuth resolution <SAR> resolution in the cross-range direction
NOTE 1 This is usually measured in terms of the impulse response of the SAR sensor and processing system. It is a function of the size of the synthetic aperture, or alternatively the dwell time (i.e. a larger aperture results in a longer dwell time results in better resolution).
[ISO 19130-1:2018]

backscattering coefficient average radar cross section per unit area
NOTE 1 If the radar return from the illuminated area is contributed by a number of independent scattering elements, it is described by the backscattering coefficient instead of radar cross section used for the point target. It

is calculated as

$\sigma^0 = \frac{\sigma}{A}$

Where

σ is the total radar cross section of an area A.

$\sigma^0$is a dimensionless parameter and is usually expressed in decibels (dB) as follows:

${\sigma^0}_{dB} = 10 \log_{10} \sigma^0$

NOTE 2 "Backscattering coefficient" is sometimes called "normalized radar cross section".

[ISO 19159-3:2018]

band range of wavelengths of electromagnetic radiation that produce a single response by a sensing device.

[ISO 19101-2:2018]

bare earth elevation height of the natural terrain free from vegetation as well as buildings and other man-made structures

[ISO 19159-2:2016]

barycentric coordinates <coordinate geometry> point in a n-dimension coordinate system using n + 1 numbers, $[u_0, u_1, u_2, u_3, \ldots, u_\mathrm{n}], 0 \leq u_i \leq 1, \sum u_\mathrm{i} = 1.0$, in which the location of a point of an n-simplex (of any dimension) is specified by a weighted centre of mass of equal masses placed at its vertices using vector algebra of the $\mathbb{R}^\mathrm{n}$ used in the coordinate reference system

NOTE 1 Even though there are n + 1 coordinates in a barycentric coordinate system, the topological dimension is n, since the restriction (sums to 1.0) loses 1 degree of freedom (once you have n ordinates, the remaining one is determined such as $u_\mathrm{n} = 1.0 - \sum_{i=0}^{\mathrm{n}-1} u_i$. The coordinates for the simplex are all non-negative, but the system can be extended outside of the simples by using negative numbers. If the ordinates are all positive, then the point is inside (interior to) the n-simplex. If one of them is 1.0 and the other 0, this is a corner of the simplex. If one of them is zero and the others still each greater than or equal to zero, the point is on the n-1-simplex opposite the vertex zeroed out. If any are negative, the point is outside of the simplex. The coordinates are dependent on the underlying coordinate reference system of the source data.

[ISO 19107:2019]

base representation <moving features> representation, using a local origin and local ordinate vectors, of a geometric object at a given reference time

NOTE 1 A rigid geometric object may undergo translation or rotation, but remains congruent with its base representation.

NOTE 2 The local origin and ordinate vectors establish an engineering coordinate reference system (ISO 19111), also called a local frame or a local Euclidean coordinate system.

[ISO 19141:2008]

base standard ISO geographic information standard or other information technology standard that is used as a source from which a profile may be constructed

[ISO 19106:2004]

basic administrative unit baunit

administrative entity, subject to registration (by law), or recordation [by informal right, or customary right, or another social tenure relationship], consisting of zero or more spatial units against which (one or more) unique and homogeneous rights [e.g. ownership right or land use right], responsibilities or restrictions are associated to the whole entity, as included in a land administration system

EXAMPLE A condominium unit comprising two spatial units (e.g. an apartment and a garage), a farm lot comprising one spatial unit (e.g. parcel of land), a servitude comprising one spatial unit (e.g. the road representing the right-of-way), a land consolidation area, or a right-of-use unit with several right holders and restricted objects.

NOTE 1 'Unique' means that a right, restriction, or responsibility is held by one or more parties (e.g. owners or users) for the whole basic administrative unit. 'Homogeneous' means that a right, restriction or responsibility (e.g. ownership, use, social tenure, lease, or easement) affects the whole basic administrative unit. For a restriction, zero parties are a possibility.

NOTE 2 A basic administrative unit may play the role of party, e.g. when the right holder is a basic administrative unit (and not a person or organization).

NOTE 3 A baunit should get a unique identifier when registered, or recorded.

NOTE 4 A baunit can consist of zero spatial units, when a registry exists, and not a cadastre.

NOTE 5 Restrictions and responsibilities can be associated with their own baunits, each with their own type of spatial unit.

[ISO 19152:2012]

basic service service providing a basic function to other services or applications in a functional manner

NOTE 1 cf. interoperate. Basic services lack any persistent, user-specific state information between invocations and are not meant for direct access by users. Because they act in a functional manner, they are readily replaceable at runtime by other services using the same interfaces.

[ISO 19132:2007]

basic test initial capability test intended to identify clear cases of non-conformance

[ISO 19105:2000]

beam width <SAR> useful angular width of the beam of electromagnetic energy

NOTE 1 Beam width is usually measured in radians and as the angular width between two points that have

50% of the power (3 dB below) of the centre of the beam. It is a property of the antenna. Power emitted outside of this angle is too little to provide a usable return.
[ISO 19130-1:2018]

bearing horizontal angle at a point relative to a specified direction
NOTE 1 The direction is usually specified to be north. In some communities the term bearing refers specifically to grid north and directions relative to true north are then termed 'azimuth'; in other communities a bearing refers specifically to true north. In this International Standard bearing is used for any specified reference direction. The angle may be reckoned positive clockwise or positive counter-clockwise depending upon the application.
[ISO 19162:2019]

bearing horizontal angle, tangent or direction at a point
NOTE 1 This definition (as opposed to the one in ISO 19162:2015) is required for this document because the concept is used in other definitions, such as first geodetic problem and second geodetic problem. The two definitions are nearly equivalent because the tangent of a curve on a surface is a tangent to the surface and does specify a direction. Usual 2D measure of bearing can be an angle equivalently measured from North clockwise, or a unit tangent vector. If the coordinate system is spatially 3D, the horizontal bearing angle may also need to a vertical altitude angle to be complete. If a reference curve (as used in ISO 19162) is parameterized by arc length, then the "derivative" is a unit vector. If another parameterization "t" is used, then the derivative should be normalized ($\vec{\tau}/\left\|\vec{\tau}\right\|; \dot{c}(t) = \vec{\tau};$). This is useful, since parameterization by arc length can be computationally difficult. The numeric representation of a vector depends on the coordinate system. The bearing is not dependent on a coordinate system, but it can be represented in any reasonable system. The bearing is not dependent on its various representations.
[ISO 19107:2019]

bicontinuous <mathematics> invertible, continuous and with a continuous inverse
[ISO 19107:2019]

binding specification of a mapping relating the information defined in a content model (data and metadata) to the data format that carries that information
[ISO 19163-1:2016]

blooming overflow of an over-saturated signal of one pixel to the neighbouring pixel
[ISO 19159-1:2014]

boresight calibration of a lidar sensor system, equipped with an Inertial Measurement Unit (IMU) and a Global Navigation Satellite System (GNSS), to accurately determine or establish its position and orientation
NOTE 1 The position of the lidar sensor system (x, y, z) is determined with respect to the GNSS antenna. The orientation (roll, pitch, heading) of the lidar sensor system is determined with respect to straight and level flight.
[ISO 19159-2:2016]

boundary set that represents the limit of an entity
NOTE 1 Boundary is most commonly used in the context of geometry, where the set is a collection of points or a collection of objects that represent those points. In other arenas, the term is used metaphorically to describe the transition between an entity and the rest of its domain of discourse.
[ISO 19107:2019]

boundary face face that is used in the 3-dimensional representation of a boundary of a spatial unit
NOTE 1 Boundary faces are used when the implied vertical and unbounded faces of a boundary face string are not sufficient to describe 3D spatial units. Boundary faces close volumes in height (e.g. every apartment floor), or in depth (e.g. an underground parking garage), or in all other directions to form a bounded volume. The volumes represent legal space (in contrast with physical space).
[ISO 19152:2012]

boundary face string boundary forming part of the outside of a spatial unit
NOTE 1 Boundary face strings are used to represent the boundaries of spatial units by means of line strings in 2D. This 2D representation is a 2D boundary in a 2D land administration system. In a 3D land administration system it represents a series of vertical boundary faces where an unbounded volume is assumed, surrounded by boundary faces which intersect the Earth's surface (such as traditionally depicted in the cadastral map).
[ISO 19152:2012]

breakline linear feature that describes a change in the smoothness or continuity of a surface
NOTE 1 A soft breakline ensures that known z-values along a linear feature are maintained (for example, elevations along a pipeline, road centreline or drainage ditch), and ensures that linear features and polygon edges are maintained in a Triangulated Irregular Network (TIN) surface model, by enforcing the breaklines as TIN edges. They are generally synonymous with 3-D breaklines because they are depicted with series of x/y/z coordinates. Somewhat rounded ridges or the trough of a drain may be collected using soft breaklines.
NOTE 2 A hard breakline defines interruptions in surface smoothness, for example, to define streams, shorelines, dams, ridges, building footprints, and other locations with abrupt surface changes.
[ISO 19159-2:2016]

broader concept superordinate concept
concept which is either a generic concept or a comprehensive concept
[ISO 1087-1:2000, ISO 19146:2018]

broadside <SAR> direction orthogonal to the velocity vector and parallel to the plane tangent to the Earth's ellipsoid at the nadir point of the ARP
[ISO 19130-1:2018]

buffer geometric object containing all points and only those points whose distance from a specified geometric object is less than or equal to a given distance use in its construction
[ISO 19107:2019]

building unit component of building (the legal, recorded or informal space of the physical entity)
EXAMPLE An apartment, a flight of stairs, a threshold, a garage, a parking place or a laundry space.
NOTE 1 A building unit may be used for different purposes (e.g. living or commercial) or it can be under construction.
[ISO 19152:2012]

calendar discrete temporal reference system that provides a basis for defining temporal position to a resolution of one day
[ISO 19108:2002]

calendar era sequence of periods of one of the types used in a calendar, counted from a specified event
[ISO 19108:2002]

calibrated focal length distance between the perspective centre and the image plane that is the result of balancing positive and negative radial lens distortions during sensor calibration
[ISO 19130-1:2018]

calibration process of quantitatively defining a system's responses to known, controlled signal inputs
[CEOS WGCV, ISO 19101-2:2018]

calibration coefficient ratio of SAR image pixel power to radar cross section without considering additive noise, after the processor gain is normalized to one, and elevation antenna pattern, range and atmospheric attenuation are all corrected
[ISO 19159-3:2018]

calibration curve expression of the relation between indication and corresponding measured quantity value
NOTE 1 A calibration curve expresses a one-to-one relation that does not supply a measurement result as it bears no information about the measurement uncertainty.
[ISO/IEC Guide 99:2007, ISO 19159-1:2014]

calibration validation process of assessing the validity of parameters
NOTE 1 With respect to the general definition of validation the "calibration validation" does only refer to a small set of parameters (attribute values) such as the result of a sensor calibration.
[ISO 19159-1:2014]

candidate route any route that satisfies all constraints of the routing request with the possible exception of optimality of the cost function
NOTE 1 Navigation is the process of finding the candidate route that optimizes a chosen cost function.
[ISO 19133:2005]

capability real-world effect that a service provider is able to provide to a service consumer
[SOA-RAF, ISO 19119:2016]

capability test test designed to determine whether an IUT conforms to a particular characteristic of an International Standard as described in the test purpose
[ISO 19105:2000]

cardinality <UML> number of elements in a set
NOTE 1 Contrast with multiplicity, which is the range of possible cardinalities a set can hold.
[UML 1, ISO 19103:2015]

Cartesian coordinate system coordinate system in Euclidean space which gives the position of points relative to n mutually perpendicular straight axes all having the same unit of measure
NOTE 1 n is 2 or 3 for the purposes of this document.
NOTE 2 A Cartesian coordinate system is a specialisation of an affine coordinate system.
[ISO 19111:2019]

catalogue collection of items or an electronic or paper document that contains information about the collection of items
[ISO 10303-227:2005, ISO 19157:2013]

character member of a set of elements that is used for the representation, organization, or control of data
[ISO/IEC 2382-1:1993, ISO 19118:2011]

characteristic abstraction of a property of an object or of a set of objects
NOTE 1 Characteristics are used for describing concepts.
[ISO 1087-1:2000, ISO 19146:2018]

checkpoint check point
point in object space (ground) used to estimate the positional accuracy of a geospatial dataset against an independent source of greater accuracy
[ISO 19159-2:2016]

child address address defined relative to a parent address
[ISO 19160-1:2015]

child addressable object addressable object that is addressed relative to another addressable object
EXAMPLE 1 An apartment within an apartment building.
EXAMPLE 2 In Japan, a jukyo bango (residence number) within a gaiku (block).

EXAMPLE 3 A building within a complex of buildings. In Korea, a dong (wing or section of a building) within a group of buildings.
[ISO 19160-1:2015]

child element <XML> immediate descendant element of an element
[ISO 19136-1:2020]

citation information object containing information that directs a reader's or user's attention from one resource to another
[ISO 24619:2011, ISO 19115-1:2014]

clarification non-substantive change to a register item
NOTE 1 A non-substantive change does not change the semantics or technical meaning of the item. Clarification does not result in a change to the registration status of the register item.
[ISO 19135-1:2015]

class <OWL> set of individuals
[OWL, ISO 19150-2:2015]

class <UML> description of a set of objects that share the same attributes, operations, methods, relationships, and semantics
[UML 1, ISO 19103:2015]

classification abstract representation of real world phenomena using classifiers
[ISO 19144-1:2009]

classification system system for assigning objects to classes
[ISO 19144-1:2009]

classified object spatial object, temporal object, or spatiotemporal object assigned to a specific legend class
[ISO 19144-1:2009]

classifier definition used to assign objects to legend classes
NOTE 1 Classifiers can be defined algorithmically or according to a set of classification system specific rules.
[ISO 19144-1:2009]

classifier <UML> mechanism that describes behavioral and structural features in any combination
[UML 1, ISO 19103:2015]

client software component that can invoke an operation from a server
[ISO 19128:2005]

closure union of the interior and boundary of a topological object or geometric object
[ISO 19107:2019]

cluster collection of targets potentially heterogeneous (each satisfying a different query criteria) whose locations fall within a small neighbourhood.
[ISO 19132:2007]

coboundary set of topological primitives of higher topological dimension associated with a particular topological object, such that this topological object is in each of their boundaries
NOTE 1 If a node is on the boundary of an edge, that edge is on the coboundary of that node. Any orientation parameter associated to one of these relations would also be associated to the other. So that if the node is the end node of the edge (defined as the end of the positive directed edge), then the positive orientation of the node (defined as the positive directed node) would have the edge on its coboundary, see ISO 19107 Figure 35.
[ISO 19107:2019]

code representation of a label according to a specified scheme
[ISO 19118:2011]

codelist value domain including a code for each permissible value
[ISO 19136-1:2020]

codespace rule or authority for a code, name, term or category
EXAMPLE Examples of codespaces include dictionaries, authorities, codelists, etc.
[ISO 19136-1:2020]

complex feature feature composed of other features
[ISO 19109:2015]

complex image first-level product produced by processing SAR Phase History Data
[ISO 19130-2:2014]

complex symbol symbol composed of other symbols of different types
EXAMPLE A dashed line symbol with a point symbol repeated at an interval.
[ISO 19117:2012]

component <UML> representation of a modular part of a system that encapsulates its contents and whose manifestation is replaceable within its environment
[UML 2, ISO 19103:2015]

composite curve sequence of curves such that each curve (except the first) starts at the end point of the previous curve in the sequence
NOTE 1 A composite curve, as a set of direct positions, has all the properties of a curve.
[ISO 19136-1:2020]

composite solid connected set of solids adjoining one another along shared boundary surfaces
NOTE 1 A composite solid, as a set of direct positions, has all the properties of a solid.
[ISO 19136-1:2020]

composite surface connected set of surfaces adjoining one another along shared boundary curves
NOTE 1 A composite surface, as a set of direct positions, has all the properties of a surface.
[ISO 19136-1:2020]

composition <UML> aggregation where the composite object (whole) has responsibility for the existence and storage of the composed objects (parts)
[UML 2, ISO 19103:2015]

compound coordinate reference system coordinate reference system using at least two independent coordinate reference systems
NOTE 1 Coordinate reference systems are independent of each other if coordinate values in one cannot be converted or transformed into coordinate values in the other.
[ISO 19111:2019]

compound registry registry containing multiple registers that share the same item classes and coordinated management of a common characteristic
NOTE 1 The common characteristic may be a shared namespace for the assignment of names and/or codes.
[ISO 19126:2009]

compound symbol symbol composed of other symbols of the same type
EXAMPLE A point symbol that is composed of two point graphics.
[ISO 19117:2012]

compression technique used for the reduction of space used by data
[ISO 19145:2013]

compression service service that accomplishes compression
[ISO 19145:2013]

computational viewpoint viewpoint on an ODP system and its environment that enables distribution through functional decomposition of the system into objects which interact at interfaces
[ISO/IEC 10746-3:2015, ISO 19154:2014]

concatenated operation coordinate operation consisting of sequential application of multiple coordinate operations
[ISO 19111:2019]

concept unit of knowledge created by a unique combination of characteristics
NOTE 1 Concepts are not necessarily bound to particular languages. They are, however, influenced by the social or cultural background which often leads to different categorizations.
[ISO 1087-1:2000, ISO 19104:2016]

concept field unstructured set of thematically related concepts
[ISO 1087-1:2000, ISO 19104:2016]

concept harmonization activity leading to the establishment of a correspondence between two or more closely related or overlapping concepts having professional, technical, scientific, social, economic, linguistic, cultural or other differences, in order to eliminate or reduce minor differences between them
NOTE 1 The purpose of concept harmonisation is to improve communication
[ISO 860:2007, ISO 19104:2016]

concept system set of concepts structured according to the relations among them
[ISO 1087-1:2000, ISO 19104:2016]

conceptual formalism set of modelling concepts used to describe a conceptual model
EXAMPLE UML meta model, EXPRESS meta model.
NOTE 1 One conceptual formalism can be expressed in several conceptual schema languages.
[ISO 19101-1:2014]

conceptual model model that defines concepts of a universe of discourse
[ISO 19101-1:2014]

conceptual schema formal description of a conceptual model
[ISO 19101-1:2014]

conceptual schema language formal language based on a conceptual formalism for the purpose of representing conceptual schemas
EXAMPLE UML, EXPRESS, IDEF1X
NOTE 1 A conceptual schema language may be lexical or graphical. Several conceptual schema languages can be based on the same conceptual formalism.
[ISO 19101-1:2014]

conditional feature portrayal function function that maps a geographic feature to a symbol based on some condition evaluated against a property or attribute of a feature
[ISO 19117:2012]

conformal angle-preserving
NOTE 1 Some projections are conformal. For example, a Mercator preserves the angle between curves, so that if two curves in a Mercator projected plane cross at a 90°, then the preimage curves on the ellipsoid also cross at 90°, such as lines of constant latitude and lines of constant longitude.
[ISO 19107:2019]

conformance fulfilment of specified requirements
[ISO 19105:2000]

conformance assessment process process for assessing the conformance of an implementation to an International Standard
[ISO 19105:2000]

conformance clause clause defining what is necessary in order to meet the requirements of the International Standard
[ISO 19105:2000]

conformance quality level threshold value or set of threshold values for data quality results used to determine how well a dataset meets the criteria set forth in its data product specification or user requirements
[ISO 19157:2013]

conformance test report summary of the conformance to the International Standard as well as all the details of the testing that supports the given overall summary
[ISO 19105:2000]

conformance testing testing of a product to determine the extent to which the product is a conforming implementation
[ISO 19105:2000]

conforming implementation implementation which satisfies the requirements
[ISO 19105:2000]

connected property of a topological space implying that only the entire space or the empty set are the only subsets which are both open and closed
NOTE 1 The formal definition of connected is that any pair of locally open sets whose union is the entire space must have a non-empty intersection.
a topological space T is connected if and only if

$$[X, Y \text{ are open}] \Rightarrow [X, Y \subset T, X \cup Y = T \Rightarrow X \cap Y \neq \varnothing]$$

This formal definition is difficult to use. The term path connected, defined below is equivalent for the purposed of this document. The use or "finite precision" coordinates makes sets which are connected but not path connected impossible to represent. In all cases "connected" is used, but "path connected" is easier to test and to visualize.
[ISO 19107:2019]

connected node node that starts or ends one or more edges
[ISO 19107:2019]

constraint restriction on how a link or turn may be traversed by a vehicle, such as vehicle classification, physical or temporal constraint
[ISO 19133:2005]

constraint <UML> condition or restriction expressed in natural language text or in a machine readable language for the purpose of declaring some of the semantics of an element
[UML 2, ISO 19103:2015]

content information set of information that is the original target of preservation or that includes part or all of that information
NOTE 1 Content information is an information object composed of its content data object and its representation information.
[ISO 14721:2012, ISO 19165-1:2018]

content model information view of an application schemas
NOTE 1 The term "information view" comes from the ISO Reference Model for Open Distributed Processing (RM-ODP) as specified in 19101-2.
[ISO 19129:2009]

context aspects or properties of an entity that affect the behavior or expectations of that entity in any given situation
[ISO 19154:2014]

context-awareness integrated operations to collect and deliver context specific information, and convert it to tailored data for each user
EXAMPLE getContext(staticFeature) = FD_Feature
NOTE 1 In the example, the getContext operation would extract geographically explicit context information FD_Feature from varied information sources, such as photos, videos, etc.
[ISO 19154:2014]

continuous change change in an attribute whose type has a distance measure such that its value can be assumed to take on intermediate values between two known measurements
NOTE 1 The interpolation of continuous change is usually done by taking into consideration constraints on the "curve" joining the two data points (time1, value1) and (time2, value2), looking at the value as a function of time. For example, if the continuous change is for the motion of a vehicle, then the constraints of physics and of the paths appropriate for that vehicle must be taken into consideration.
[ISO 19132:2007]

continuous coverage coverage that returns different values for the same feature attribute at different direct positions within a single spatial object, temporal object, or spatiotemporal object in its domain
NOTE 1 Although the domain of a continuous coverage is ordinarily bounded in terms of its spatial and/or temporal extent, it can be subdivided into an infinite number of direct positions.
[ISO 19123:2005]

control body group of technical experts that makes decisions regarding the content of a register
[ISO 19135-1:2015]

control point <coordinate geometry> point used in the construction of a geometry that partially controls its shape but does not necessarily lie on the geometry
NOTE 1 A centre of an arc is a control point; poles in b-spline curves are control points.
[ISO 19107:2019]

conversion transformation from one format to another
[ISO 19145:2013]

conversion rule rule for converting instances in the input data structure to instances in the output data structure
[ISO 19118:2011]

conversion service service that invokes a converter
[ISO 19145:2013]

converter resource that performs conversion
NOTE 1 The resource can be a device or software.
[ISO 19145:2013]

convex <geometry> containing all points on a "line" joining two interior points

NOTE 1 The definition of convex requires a definition of line. For coordinate systems, this is the usual linear interpolated arc, but in context, the "line" on a geometric reference surface will be a "geodesic arc". The default in this document is the linear interpolate.
[ISO 19107:2019]

convex hull smallest convex set containing a given geometric object
NOTE 1 "Smallest" is the set theoretic smallest, not an indication of a measurement. The definition can be rewritten as "the intersection of all convex sets that contain the geometric object". Another definition in a Euclidean space $\mathbb{E}^n$ is the union of all lines with both end points in the given geometric object.

$$C = A.\text{convexHull} \Leftrightarrow$$
$$[C.\text{convex} = \text{TRUE}] \wedge [A \subset C]$$
$$\wedge[[B.\text{convex} = \text{true}, A \subset B] \Rightarrow [A \subseteq C \subseteq B]]$$

[ISO 19107:2019]

convex set geometric set in which any direct position on the straight-line segment joining any two direct positions in the geometric set is also contained in the geometric set
NOTE 1 Convex sets are "simply connected", meaning that they have no interior holes, and can normally be considered topologically isomorphic to a Euclidean ball of the appropriate dimension. So the surface of a sphere can be considered to be geodesically convex.
[Dictionary of Computing:1996, ISO 19125:2005]

coordinate one of a sequence of numbers designating the position of a point
NOTE 1 In a spatial coordinate reference system, the coordinate numbers are qualified by units.
[ISO 19111:2019]

coordinate conversion coordinate operation that changes coordinates in a source coordinate reference system to coordinates in a target coordinate reference system in which both coordinate reference systems are based on the same datum
EXAMPLE 1 A mapping of ellipsoidal coordinates to Cartesian coordinates using a map projection.
EXAMPLE 2 Change of units such as from radians to degrees or from feet to metres.
NOTE 1 A coordinate conversion uses parameters which have specified values.
[ISO 19111:2019]

coordinate dimension <coordinate geometry> number of separate decisions needed to describe a position in a coordinate system
NOTE 1 The coordinate dimension represents the number of choices made, and constraints can restrict choices. A barycentric coordinate which has (n + 1)-offsets, but the underlying space is dimension n. Homogeneous coordinates (wx, wy, wz, w) are actually 3 dimensional because the choice of "w" does not affect the position, i.e. $(wx, wy, wz, w) = (x, y, z, 1) \rightarrow (x, y, z)$ which is not affected by w. The dimension will be at most the count of the numbers in the coordinate, but it can be less if the coordinates are constrained in some manner.
[ISO 19107:2019]

coordinate epoch epoch to which coordinates in a dynamic coordinate reference system are referenced
[ISO 19111:2019]

coordinate operation process using a mathematical model, based on a one-to-one relationship, that changes coordinates in a source coordinate reference system to coordinates in a target coordinate reference system, or that changes coordinates at a source coordinate epoch to coordinates at a target coordinate epoch within the same coordinate reference system
[ISO 19111:2019]

coordinate reference system coordinate system that is related to an object by a datum
NOTE 1 For geodetic and vertical datums, the object will be the Earth.
[ISO 19111:2019]

coordinate set collection of coordinate tuples referenced to the same coordinate reference system and if that coordinate reference system is dynamic also to the same coordinate epoch
[ISO 19111:2019]

coordinate system set of mathematical rules for specifying how coordinates are to be assigned to points
[ISO 19111:2019]

coordinate transformation coordinate operation that changes coordinates in a source coordinate reference system to coordinates in a target coordinate reference system in which the source and target coordinate reference systems are based on different datums
NOTE 1 A coordinate transformation uses parameters which are derived empirically. Any error in those coordinates will be embedded in the coordinate transformation and when the coordinate transformation is applied the embedded errors are transmitted to output coordinates.
NOTE 2 A coordinate transformation is colloquially sometimes referred to as a 'datum transformation'. This is erroneous. A coordinate transformation changes coordinate values. It does not change the definition of the datum. In this document coordinates are referenced to a coordinate reference system. A coordinate transformation operates between two coordinate reference systems, not between two datums.
[ISO 19111:2019]

coordinate tuple tuple composed of coordinates
NOTE 1 The number of coordinates in the coordinate tuple equals the dimension of the coordinate system; the order of coordinates in the coordinate tuple is identical to the order of the axes of the coordinate system.
[ISO 19111:2019]

Coordinated Universal Time (UTC) time scale maintained by the Bureau International des Poids et Mesures (International Bureau of Weights and Measures) and the International Earth Rotation Service (IERS) that forms the basis of a coordinated dissemination of standard frequencies and time
[ITU-R Rec.TF.686-1 (1997), ISO 19108:2002]

correction compensation for an estimated systematic effect
NOTE 1 See ISO/IEC Guide 98-3:2008, 3.2.3, for an explanation of "systematic effect".
NOTE 2 The compensation can take different forms, such as an addend or a factor, or can be deduced from a table.
[ISO/IEC Guide 99:2007, ISO 19159-1:2014]

correctness correspondence with the universe of discourse
[ISO 19157:2013]

Correspondence Model functional relationship between ground and image coordinates based on the correlation between a set of ground control points and their corresponding image coordinates
[ISO 19130-1:2018]

cost function function that associates a measure (cost) to a route
NOTE 1 The normal mechanism is to apply a cost to each part of a route, and to define the total route cost as the sum of the cost of the parts. This is necessary for the operation of the most common navigation algorithms. The units of cost functions are not limited to monetary costs and values only, but include such measures as time, distance, and possibly others. The only requirement is that the function be additive and at least non-negative. This last criteria can be softened as long as no zero or less cost is associated with any loop in the network, as this will prevent the existence of a "minimal cost" route.
[ISO 19133:2005]

coupling linkage of two or more software systems through information transfer or messaging
NOTE 1 Compare with integration. While the conceptual schema of the information transferred shall be agreed upon to some level, coupling applications can be and are usually flexible in the data representation of that information as long as the semantics content is correct and mappable to some canonical representation of the conceptual schema. The most common mapping technology used for XML messages is XSLT, and the transformation stylesheet can be supplied either by the service broker or by the service provider. It is considered a best practice for a service provider to supply his functionality through several logically equivalent messaging APIs, each represented by a different URI linked to an XSLT transformation bridge, and implemented by the same internal code.
NOTE 2 Loose coupling and tight coupling are not at present well-defined terms in the literature. Generally, "tight" coupling means that there is some sort of incurred dependency between requester and responder in the use of the interface, while "loose" means no such dependency. The nature of that dependency is not consistently defined between authors. In that light, "tight" coupling or "tight" integration are both bad practices, and have been viewed as such since the inception of the terms. Some literature refers to integration as "tight coupling", but that is a less accurate description.
[ISO 19132:2007]

coverage feature that acts as a function to return values from its range for any direct position within its spatial, temporal or spatiotemporal domain
EXAMPLE Examples include a raster image, polygon overlay, or digital elevation matrix.
NOTE 1 In other words, a coverage is a feature that has multiple values for each attribute type, where each direct position within the geometric representation of the feature has a single value for each attribute type.
[ISO 19123:2005]

coverage geometry configuration of the domain of a coverage described in terms of coordinates
[ISO 19123:2005]

cross-map entry part of a cross-mapping data collection which documents the cross-mapped relationships between two concepts
[ISO 19146:2018]

cross-map register register of cross-map entries
NOTE 1 A cross-map register may be realized as a subregister in a hierarchical register. In such cases, the term "cross-map subregister" may be used.
[ISO 19146:2018]

cross-mapping comparison of terminological entries from different domains to determine their semantic relationship
[ISO 19146:2018]

cross-talk any signal or circuit unintentionally affecting another signal or circuit
NOTE 1 For PolSAR sensor, if the transmitting channel is horizontally (H) polarized, the cross-talk on transmitting defines the ratio of V polarization transmitting power to H polarization transmitting power, expressed in decibels (dB). The cross-talk on receiving is similar to that on transmitting.
[ISO 19159-3:2018]

cross-track perpendicular to the direction in which the collection platform moves
[ISO 19130-1:2018]

curvature vector <differential geometry> second derivative of a curve parameterized by arc length, at a point
NOTE 1 If $c(s) = (x(s), y(s), z(s))$ is a curve in 3D Cartesian space ($\mathbb{E}^3$), and s is the arc length along $c(s)$, then the unit tangent vector is $\dot{c}(s) = (\dot{x}(s), \dot{y}(s), \dot{z}(s))$, i.e. the derivative of the coordinate values of "c" with respect to "s". The curvature vector is $\ddot{c}(s) = (\ddot{x}(s), \ddot{y}(s), \ddot{z}(s))$. The curvature vector can be approximated by the inverse of the radius of a circle through any 3 nearby points on the curve (pointed from the curve to towards the centre of the circle)
[ISO 19107:2019]

curve 1-dimensional geometric primitive, representing the continuous image of a line
NOTE 1 The boundary of a curve is the set of points at either end of the curve. If the curve is a cycle, the two ends are identical, and the curve (if topologically closed) is considered to not have a boundary. The first point is called the start point, and the last is the end point. Connectivity of the curve is guaranteed by the "continuous image of a line" clause. A topological theorem states that a continuous image of a connected set is connected.
[ISO 19136-1:2020]

customer organization or person that receives a product
NOTE 1 The customer can be internal or external to the supplier organisation.
[ISO 9000:2005, ISO 19158:2012]

cycle <geometry, topology> bounded spatial object with an empty boundary
NOTE 1 Cycles are used to describe boundary components. A cycle usually has no boundary because it closes on itself, but it is bounded (i.e., it does not have infinite extent). A circle or a sphere, for example, has no boundary (i.e., its boundary is empty), but is bounded.
[ISO 19107:2019]

cylindrical coordinate system three-dimensional coordinate system in Euclidean space in which position is specified by two linear coordinates and one angular coordinate
[ISO 19111:2019]

dark current output current of a photoelectric detector (or of its cathode) in the absence of incident radiation
NOTE 1 For calibration of optical sensors dark current is measured by the absence of incident optical radiation.
[ISO 19159-1:2014]

dark current noise noise of current at the output of a detector, when no optical radiation is sensed
[ISO 19159-1:2014]

dark signal non uniformity DSNU
response of a detector element if no visible or infrared light is present
NOTE 1 This activation is mostly caused by imperfection of the detector.
[ISO 19159-1:2014]

data reinterpretable representation of information in a formalised manner suitable for communication, interpretation, or processing
[ISO/IEC 2382-1:1993, ISO 19118:2011]

data category result of the specification of a specific type of terminological data
[ISO 10241-1:2011, ISO 19104:2016]

data compaction reduction of the number of data elements, bandwidth, cost, and time for the generation, transmission, and storage of data without loss of information by eliminating unnecessary redundancy, removing irrelevancy, or using special coding
NOTE 1 Whereas data compaction reduces the amount of data used to represent a given amount of information, data compression does not.
NOTE 2 Data compaction can be done through aggregation of like values in adjacent grid cells, tiling schemes or other means of eliminating information that is not relevant.
[ANSI T1.523-2001, ISO 19129:2009]

data compression reducing either the amount of storage space required to store a given amount of data, or the length of message required to transfer a given amount of information
[ISO 19129:2009]

data dictionary formal repository of terms used to describe data
[ISO 14721:2012, ISO 19165-1:2018]

data dissemination session delivery of media or a single telecommunications session that provides data to a consumer
NOTE 1 The data dissemination session format/contents is based on a data model negotiated between the OAIS and the consumer in the request agreement. This data model identifies the logical constructs used by the OAIS and how they are represented on each media delivery or in the telecommunication session.
[ISO 14721:2012, ISO 19165-1:2018]

data interchange delivery, receipt and interpretation of data
[ISO 19118:2011]

data point <coordinate geometry> point that lies on the geometry
NOTE 1 The vertices in a line string are data points, the points used to construct a polynomial spline are data points. Data points can be used as control points, but are often derived after the geometry is constructed.
[ISO 19107:2019]

data product dataset or dataset series that conforms to a data product specification
[ISO 19131:2007]

data product specification detailed description of a dataset or dataset series together with additional information that will enable it to be created, supplied to and used by another party
NOTE 1 A data product specification provides a description of the universe of discourse and a specification for mapping the universe of discourse to a dataset. It may be used for production, sales, end-use or other purposes.
[ISO 19131:2007]

data property <OWL> semantic association between an individual and a typed literal
NOTE 1 Data properties were sometimes referred to as 'concrete properties' in Description Logic.
[OWL, ISO 19150-2:2015]

data quality basic measure generic data quality measure used as a basis for the creation of specific data quality measures
NOTE 1 Data quality basic measures are abstract data types. They cannot be used directly when reporting data quality.
[ISO 19157:2013]

data submission session delivery of media or a single telecommunications session that provides data to an OAIS
NOTE 1 The data submission session format/contents is based on a data model negotiated between the OAIS and the producer in the submission agreement. This data model identifies the logical constructs used by the producer and how they are represented on each media delivery or in the telecommunication session.
[ISO 14721:2012, ISO 19165-1:2018]

data transfer movement of data from one point to another over a medium
NOTE 1 Transfer of information implies transfer of data.
[ISO 19118:2011]

data type specification of a value domain with operations allowed on values in this domain
EXAMPLE Integer, Real, Boolean, String and Date.
NOTE 1 Data types include primitive predefined types and user-definable types.
[ISO 19103:2015]

dataset identifiable collection of data
NOTE 1 A dataset may be a smaller grouping of data which, though limited by some constraint such as spatial extent or feature type, is located physically within a larger dataset. Theoretically, a dataset may be as small as a single feature or feature attribute contained within a larger dataset. A hardcopy map or chart may be considered a dataset.
[ISO 19115-1:2014]

dataset series collection of datasets sharing common characteristics
[ISO 19115-1:2014]

datatype <OWL> entities that refer to a set of concrete data values
EXAMPLE xsd:string, xsd:integer, xsd:decimal
NOTE 1 Datatypes are distinct from classes of individuals, the latter are denoted by URIs and may be used by reference.
[OWL, ISO 19150-2:2015]

datum reference frame
parameter or set of parameters that realize the position of the origin, the scale, and the orientation of a coordinate system
[ISO 19111:2019]

datum ensemble group of multiple realizations of the same terrestrial or vertical reference system that, for approximate spatial referencing purposes, are not significantly different
EXAMPLE WGS 84 as an undifferentiated group of realizations including WGS 84 (TRANSIT), WGS 84 (G730), WGS 84 (G873), WGS 84 (G1150), WGS 84 (G1674) and WGS 84 (G1762). At the surface of the Earth these have changed on average by 0.7 m between the TRANSIT and G730 realizations, a further 0.2 m between G730 and G873, 0.06 m between G873 and G1150, 0.2 m between G1150 and G1674 and 0.02 m between G1674 and G1762).
NOTE 1 Datasets referenced to the different realizations within a datum ensemble may be merged without coordinate transformation.
NOTE 2 'Approximate' is for users to define and typically is in the order of under 1 decimetre but may be up to 2 metres.
[ISO 19111:2019]

day period having a duration nominally equivalent to the periodic time of the Earth's rotation around its axis
[ISO 19108:2002]

definition representation of a concept by a descriptive statement which serves to differentiate it from related concepts
[ISO 1087-1:2000, ISO 19104:2016]

Delaunay triangulation network of triangles such that the circle passing through the vertices of any triangle does not contain, in its interior, the vertex of any other triangle
[ISO 19123:2005]

delimiting characteristic essential characteristic used for distinguishing a concept from related concepts
NOTE 1 The delimiting characteristic support for the back may be used for distinguishing the concepts "stool" and "chair".
[ISO 1087-1:2000, ISO 19146:2018]

delivery <postal> process in which a postal item leaves the responsibility of the postal operator through being handed over to, or left for collection by, the addressee, the mailee

or an authorized representative, or deposited in a private letter box accessible to one or other of these
NOTE 1 Delivery does not always imply receipt by the addressee or mailee.
[ISO 19160-4:2017]

delivery address <postal> postal address which the postal operator is requested to use to deliver the postal item
NOTE 1 In the normal case, the delivery address is the same as the postal address specified by the mailer.
NOTE 2 The delivery address may in certain circumstances, e.g. unaddressed mail, not actually be represented on the postal item. In this case, the delivery address is determined by the postal operator in accordance with an agreement between the operator and the mailer.
NOTE 3 The postal item might not actually be delivered to the requested delivery address. For example, in the case of forwarding, delivery takes place at the forwarding address.
[ISO 19160-4:2017]

delivery point <postal> physical location recognized by a postal operator as a valid location at which delivery may occur
[ISO 19160-4:2017]

dependency <UML> relationship that signifies that a single or a set of model elements requires other model elements for their specification or implementation
NOTE 1 This means that the complete semantics of the depending elements is either semantically or structurally dependent on the definition of the supplier element(s).
[UML 2, ISO 19103:2015]

deprecated term term rated according to the scale of the term acceptability rating as undesired
[ISO 1087-1:2000, ISO 19104:2016]

depression angle vertical angle from the platform horizontal plane to the slant range direction, usually measured at the ARP
NOTE 1 Approximately the complement of the look angle
[ISO 19130-2:2014]

depth distance of a point from a chosen vertical reference surface downward along a line that is perpendicular to that surface
NOTE 1 The line direction may be straight, or be dependent on the Earth's gravity field or other physical phenomena.
NOTE 2 A depth above the vertical reference surface will have a negative value.
[ISO 19111:2019]

derived coordinate reference system coordinate reference system that is defined through the application of a specified coordinate conversion to the coordinates within a previously established coordinate reference system
NOTE 1 The previously established coordinate reference system is referred to as the base coordinate reference system.
NOTE 2 A derived coordinate reference system inherits its datum or reference frame from its base coordinate reference system.
NOTE 3 The coordinate conversion between the base and derived coordinate reference system is implemented using the parameters and formula(s) specified in the definition of the coordinate conversion.
[ISO 19111:2019]

design coordinate reference system engineering coordinate reference system in which the base representation of a moving object is specified
[ISO 19141:2008]

designated community identified group of potential consumers who should be able to understand a particular set of information
NOTE 1 The designated community may be composed of multiple user communities. A designated community is defined by the archive and this definition may change over time.
[ISO 14721:2012, ISO 19165-1:2018]

designation designator
representation of a concept by a sign which denotes it
NOTE 1 In terminology work, three types of designations are distinguished: symbols, appellations and terms.
[ISO 1087-1:2000, ISO 19104:2016]

detector device that generates an output signal in response to an energy input
[ISO 19130-1:2018]

deviation divergence from a plan or the normal situation
NOTE 1 These may be deviations with respect to the time schedule or the accomplishment of the transport or deviations with respect to the services or facilities that are provided.
[ISO 19147:2015]

diameter <metric> maximum distance between two points in the set of points
[ISO 19107:2019]

Differential Global Navigational Satellite System enhancement to Global Positioning System that uses GNSS and DGNSS to broadcast the difference between the positions indicated by the satellite systems and the known fixed positions
[ISO 19130-2:2014]

digital elevation model dataset of elevation values that are assigned algorithmically to 2-dimensional coordinates
[ISO 19101-2:2018]

digital item structured digital object [asset, work, service, data or information] with a standard representation, identification and metadata framework
[ISO/IEC TR 21000-1:2004, ISO 19132:2007]

digital migration transfer of digital information, while intending to preserve it, within the OAIS
NOTE 1 Digital migration is distinguished from transfers in general by three attributes:
– a focus on the preservation of the full information content that needs preservation;
– a perspective that the new archival implementation of the information is a replacement for the old;
– an understanding that full control and responsibility over all aspects of the transfer resides with the OAIS.
[ISO 14721:2012, ISO 19165-1:2018]

digital number DN
integer value representing a measurement as detected by a sensor
[ISO 19101-2:2018]

digital object object composed of a set of bit sequences
[ISO 14721:2012, ISO 19165-1:2018]

digital surface model DSM
digital elevation model (DEM) that depicts the elevations of the top surfaces of buildings, trees, towers, and other features elevated above the bare earth
NOTE 1 DSMs are especially relevant for telecommunications management, air safety, forest management, and 3-D modelling and simulation.
[ISO 19159-2:2016]

digital terrain model DTM
digital elevation model (DEM) that incorporates the elevation of important topographic features on the land.
NOTE 1 DTMs are comprised of mass points and breaklines that are irregularly spaced to better characterize the true shape of the bare-earth terrain. The net result of DTMs is that the distinctive terrain features are more clearly defined and precisely located, and contours generated from DTMs more closely approximate the real shape of the terrain.
[ISO 19159-2:2016]

Dijkstra graph positively weighted directed graph appropriately configured to execute a shortest path search
NOTE 1 The term comes from the most commonly known algorithm for finding a shortest path in a positively weighted graph. Although this algorithm is not the only one in use, the requirements for the graph are common to most. The most common relaxation of the requirement is the "positive weights", which are not needed in the Bellman-Ford algorithm.
[ISO 19133:2005]

direct evaluation method method of evaluating the quality of a dataset based on inspection of the items within the dataset
[ISO 19157:2013]

direct position position described by a single set of coordinates within a coordinate reference system
[ISO 19136-1:2020]

directed edge directed topological object that represents an association between an edge and one of its orientations
NOTE 1 A directed edge that is in agreement with the orientation of the edge has a + orientation, otherwise, it has the opposite (−) orientation. Directed edge is used in topology to distinguish the right side (−) from the left side (+) of the same edge and the start node (−) and end node (+) of the same edge and in computational topology to represent these concepts.
[ISO 19107:2019]

directed face directed topological object that represents an association between a face and one of its orientations
NOTE 1 The orientation of the directed edges that compose the exterior boundary of a directed face will appear positive from the direction of this vector; the orientation of a directed face that bounds a topological solid will point away from the topological solid. Adjacent solids would use different orientations for their shared boundary, consistent with the same sort of association between adjacent faces and their shared edges. Directed faces are used in the coboundary relation to maintain the spatial association between face and edge.
[ISO 19107:2019]

directed node directed topological object that represents an association between a node and one of its orientations
NOTE 1 Directed nodes are used in the coboundary relation to maintain the spatial association between edge and node. The orientation of a node is with respect to an edge, "+" for end node, "−" for start node. This is consistent with the vector notion of "result = end − start".
[ISO 19107:2019]

directed solid directed topological object that represents an association between a topological solid and one of its orientations
NOTE 1 Directed solids are used in the coboundary relation to maintain the spatial association between face and topological solid. The orientation of a solid is with respect to a face, "+" if the upNormal is outward, "−" if inward. This is consistent with the concept of "up = outward" for a surface bounding a solid.
[ISO 19107:2019]

discrete change change in an attribute value such that it can be assumed to have changed without having taken intermediate values between two known measurements
NOTE 1 Legal changes of parcel changes are discrete, having occurred at a specific time.
[ISO 19132:2007]

discrete coverage coverage that returns the same feature attribute values for every direct position within any single spatial object, temporal object, or spatiotemporal object in its domain

NOTE 1 The domain of a discrete coverage consists of a finite set of spatial, temporal, or spatiotemporal objects.
[ISO 19123:2005]

discrete spatiotemporal object temporal sequence of object representations depicting the same spatial feature at different times
[ISO 19132:2007]

dissemination information package DIP
information package, derived from one or more AIPs, and sent by archives to the consumer in response to a request to the OAIS
[ISO 14721:2012, ISO 19165-1:2018]

distance <geometry, metric spaces> minimal length of a curve that joins the two points or geometries
NOTE 1 The usual distance function for two points in a coordinate space assumes an underlying plane and is a Euclidean distance. If the underlying Reference Surface is not a plane, then distance is defined by this minimum length of all curves between the two points. These surfaces are prime examples of non-Euclidean geometry, where the parallel postulate in Euclid's Elements does not hold. In mathematical terms, distance is the "greatest lower bound" of the length of the curves. The word minimum is sometimes used, but there should be no expectation that an instance of that minimum actually occurs, only that any larger number will have a length in the set that is smaller.
[ISO 19107:2019]

distance measure distance metric
measure of the pairs of values of an attribute type that assigns a numeric value that is positive, symmetric and satisfies the triangular inequality
NOTE 1 A measure "d" is positive if $d(x, y) > 0$ for every x, y where $x \neq y$ and $d(x, x) = 0$. A measure "d" is symmetric if $d(x, y) = d(y, x)$ for every x, y. A measure "d" satisfies the triangular inequality if $d(x, y) \leqq d(x, a) + d(a, y)$ for every a, x and y. All numeric or vector valued attributes have such a metric, the most common being the Euclidean metric based on the square root of the sum of the squares of the differences in each dimension. Other non-Euclidean metrics take "curvature of space" into account (such as along the surface of the spheroid).
[ISO 19132:2007]

distribution transparency property of hiding from a particular user the potential behavior of some parts of a distributed system
NOTE 1 Distribution transparencies enable complexities associated with system distribution to be hidden from applications where they are irrelevant to their purpose.
[ISO/IEC 10746-2:2009, ISO 19119:2016]

document <XML> well-formed data object
[W3C XML, ISO 19157-2:2016]

domain well-defined set
NOTE 1 Domains are used to define the domain set and range set of attributes, operators and functions.
[ISO 19109:2015]

domain <general vocabulary> distinct area of human knowledge to which a terminological entry is assigned
NOTE 1 Within a database or other terminology collection, a set of domains will generally be defined. More than one domain can be associated with a given concept.
[ISO 19104:2016]

domain <ontology> restriction to constrain the subject class which participates in a subject-predicate-object triple
[ISO 19150-4:2019]

domain concept concept that is associated with a specific domain
NOTE 1 A concept may be associated with several domains and separately identified as a domain concept in relation to each.
[ISO 19146:2018]

domain feature feature of a type defined within a particular application domain
NOTE 1 This may be contrasted with observations and sampling features, which are features of types defined for cross-domain purposes.
[ISO 19156:2011]

Doppler angle <SAR> angle between the velocity vector and the range vector.
[ISO 19130-1:2018]

Doppler shift wavelength change resulting from relative motion of source and detector
NOTE 1 In the SAR context, it is the frequency shift imposed on a radar signal due to relative motion between the transmitter and the object being illuminated.
[ISO 19130-1:2018]

draught vertical distance, at any section of a vessel from the surface of the water to the bottom of the keel
[IHO Hydrographic Dictionary S-32, Fifth Edition, ISO 19130-2:2014]

dynamic conversion online and real time conversion of data
[ISO 19145:2013]

dynamic coordinate reference system coordinate reference system that has a dynamic reference frame
NOTE 1 Coordinates of points on or near the crust of the Earth that are referenced to a dynamic coordinate reference system may change with time, usually due to crustal deformations such as tectonic motion and glacial isostatic adjustment.
NOTE 2 Metadata for a dataset referenced to a dynamic coordinate reference system should include coordinate epoch information.
[ISO 19111:2019]

dynamic datum dynamic reference frame
reference frame in which the defining parameters include time evolution
NOTE 1 The defining parameters that have time evolution are usually a coordinate set.
[ISO 19111:2019]

dynamic reference frame dynamic datum
reference frame in which the defining parameters include time evolution
NOTE 1 The defining parameters that have time evolution are usually a coordinate set.
[ISO 19111:2019]

easting *E*
distance in a coordinate system, eastwards (positive) or westwards (negative) from a north-south reference line
[ISO 19111:2019]

edge 1-dimensional topological primitive
NOTE 1 The geometric realization of an edge is a curve. The boundary of an edge is the set of one or two nodes associated to the edge within a topological complex.
[ISO 19107:2019]

edge-node graph graph embedded within a topological complex composed of all of the edges and connected nodes within that complex
NOTE 1 The edge-node graph is a subcomplex of the complex within which it is embedded.
[ISO 19107:2019]

e-government digital interaction between a government and citizens, government and businesses, and between government agencies
[ISO 19101-1:2014]

element <XML> basic information item of an XML document containing child elements, attributes and character data
NOTE 1 From the XML Information Set: "Each XML document contains one or more elements, the boundaries of which are either delimited by start-tags and end-tags, or, for empty elements, by an empty-element tag. Each element has a type, identified by name, sometimes called its 'generic identifier' (GI), and may have a set of attribute specifications. Each attribute specification has a name and a value.
[ISO 19136-1:2020]

ellipsoid reference ellipsoid
<geodesy> geometric reference surface embedded in 3D Euclidean space formed by an ellipse that is rotated about a main axis
NOTE 1 For the Earth the ellipsoid is bi-axial with rotation about the polar axis. This results in an oblate ellipsoid with the midpoint of the foci located at the nominal centre of the Earth.
[ISO 19111:2019]

ellipsoid reference ellipsoid <geodesy> geometric reference surface embedded in 3D Euclidean space represented by an ellipsoid of revolution where the rotation is about the polar axis
NOTE 1 For the Earth the rotation is about the polar axis. This results in an oblate ellipsoid with midpoint of the foci located at the nominal centre of the Earth.
NOTE 2 The two usual algorithms for latitude on an ellipsoid and on a sphere (such as used in spherical coordinates) are only equivalent if the ellipsoid is a sphere, having all radii equal in all directions. The problem is that a radial line from the centre of a general ellipsoid does not always cross the surface of the ellipsoid orthogonally. In general, planar slices through the centre do not intersect the surface orthogonally, and therefore the curves that correspond to the great circles of a sphere are not geodesics on the ellipsoid.
NOTE 3 The topology of the ellipsoid is inherited from the $\mathbb{E}^3$ space in which it is embedded. The difference is that metrics such as distance and direction on the ellipsoid are restricted to curves wholly on the surface and vectors tangent to the surface.
[ISO 19107:2019]

ellipsoidal coordinate system geodetic coordinate system
coordinate system in which position is specified by geodetic latitude, geodetic longitude and (in the three-dimensional case) ellipsoidal height
[ISO 19111:2019]

ellipsoidal height geodetic height
h
distance of a point from the reference ellipsoid along the perpendicular from the reference ellipsoid to this point, positive if upwards or outside of the reference ellipsoid
NOTE 1 Only used as part of a three-dimensional ellipsoidal coordinate system or as part of a three-dimensional Cartesian coordinate system in a three-dimensional projected coordinate reference system, but never on its own.
[ISO 19111:2019]

ellipsoidal latitude geodetic latitude
φ
angle from the equatorial plane to the perpendicular to the ellipsoid through a given point, northwards treated as positive
[ISO 19111:2019]

ellipsoidal longitude geodetic longitude
λ
angle from the prime meridian plane to the meridian plane of a given point, eastward treated as positive
[ISO 19111:2019]

empty set $\varnothing$
<mathematics> set without any elements
NOTE 1 Sets are equal if they contain exactly the same elements. Since any two empty sets would share ex-

actly the same contained elements (by definition none), they are, by definition, equal. The empty set (∅) can be considered a geometric entity, because all the elements it contains are points. This is a vacuous statement since the set ∅ contains no elements, and therefore the "for all" statement has nothing to test and is thus true in each of its non-existent cases. There are a lot of true but vacuous statements in proofs about ∅. This confuses some programmers since many systems use type safe sets, in which the class of the entities determines a class for the container set. The math does not care about "class" and only sees sets; so that an empty set of aardvarks and an empty set of zebras in mathematics are (is?) the same set. The other confusion is that ∅ is not the database Null introduced by Codd and used in relational and other query languages in 3-valued logic. Null means unknown and many statements involving Null are undecidable (neither provably true nor provably false). The empty set is not "lack of knowledge" but certainty in the nonexistence of elements in the set. Most statements beginning "for all elements in ∅" are true, but vacuous. Most statements beginning "there exist an element in ∅" are always categorically false. It is almost impossible to construct an undecidable statement about ∅. Null and ∅ are not related. "Void" can mean "invalid" or "completely empty".
[ISO 19107:2019]

encoding conversion of data into a series of codes
[ISO 19118:2011]

encoding rule identifiable collection of conversion rules that define the encoding for a particular data structure
EXAMPLE XML, ISO 10303-21, ISO/IEC 8211.
NOTE 1 An encoding rule specifies the types of data to be converted as well as the syntax, structure and codes used in the resulting data structure.
[ISO 19118:2011]

encoding service software component that has an encoding rule implemented
[ISO 19118:2011]

end node <topology> node in the boundary of an edge that corresponds to the end point of that edge
[ISO 19107:2019]

end point last point of a curve
[ISO 19107:2019]

engineering coordinate reference system coordinate reference system based on an engineering datum
EXAMPLES Local engineering and architectural grids; coordinate reference system local to a ship or an orbiting spacecraft.
[ISO 19111:2019]

engineering datum local datum
datum describing the relationship of a coordinate system to a local reference
EXAMPLE A system for identifying relative positions within a few kilometres of the reference point.
NOTE 1 Engineering datum excludes both geodetic and vertical datums.
[ISO 19111:2019]

engineering viewpoint viewpoint on an ODP system and its environment that focuses on the mechanisms and functions required to support distributed interaction between objects in the system
[ISO/IEC 10746-3:2009, ISO 19119:2016]

enterprise viewpoint viewpoint on an ODP system and its environment that focuses on the purpose, scope and policies for that system
[ISO/IEC 10746-3:2009, ISO 19154:2014]

entity something that has separate and distinct existence and objective or conceptual reality
[ISO 19119:2016]

epoch <geodesy> point in time
EXAMPLE 2017-03-25 in the Gregorian calendar is epoch 2017.23.
NOTE 1 In this document an epoch is expressed in the Gregorian calendar as a decimal year.
[ISO 19111:2019]

error, measurement error, error of measurement
measured quantity value minus a reference quantity value
NOTE 1 The concept of "measurement error" can be used both
a) when there is a single reference quantity value to refer to, which occurs if a calibration is made by means of a measurement standard with a measured quantity value having a negligible measurement uncertainty or if a conventional quantity value is given, in which case the measurement error is known, and
b) if a measurand is supposed to be represented by a unique true quantity value or a set of true quantity values of negligible range, in which case the measurement error is not known.
NOTE 2 Measurement error should not be confused with production error or mistake.
[ISO/IEC Guide 99:2007, ISO 19159-2:2016]

error budget <metric> statement of or methodology for describing the nature and magnitude of the errors which affect the results of a calculation
NOTE 1 In the most usual case, error budgets in this document describe metric calculations using representational geometry objects to estimate real-world metrics, such as distance and area.
[ISO 19107:2019]

error propagation process of determining the uncertainties of derived quantities from the known uncertainties of the quantities on which the derived quantity is dependent
NOTE 1 Error propagation is governed by the mathematical function relating the derived quantity to the quantities from which it was derived.
[ISO 19130-1:2018]

essential characteristic characteristic which is indispensable to understanding a concept
[ISO 1087-1:2000, ISO 19146:2018]

evaluation <coverage> determination of the values of a coverage at a direct position within the domain of the coverage
[ISO 19123:2005]

event action which occurs at an instant
[ISO 19108:2002]

executable test case specific test of an implementation to meet particular requirements
NOTE 1 Instantiation of an abstract test case with values.
[ISO 19105:2000]

executable test suite ETS
set of executable test cases
[ISO 19105:2000]

exponential map <differential geometry> function that maps tangent vectors at a point to end point of geodesic beginning at that point with an exit bearing equal to that of the vector and a length equal to that of the vector
NOTE 1 See first geodetic problem for an explanation of the process of calculating this mapping.
[ISO 19107:2019]

ex-situ referring to the study, maintenance or conservation of a specimen or population away from its natural surroundings
NOTE 1 Opposite of in-situ.
[ISO 19156:2011]

exterior difference between the universe and the closure
NOTE 1 The concept of exterior is applicable to both topological and geometric complexes.
[ISO 19107:2019]

external accuracy absolute accuracy
closeness of reported coordinate values to values accepted as or being true
NOTE 1 Where the true coordinate value may not be perfectly known, accuracy is normally tested by comparison to available values that can best be accepted as true.
[ISO 19159-2:2016]

external coordinate reference system coordinate reference system whose datum is independent of the object that is located by it
[ISO 19130-1:2018]

face 2-dimensional topological primitive
NOTE 1 The geometric realization of a face is a surface. The boundary of a face is the set of directed edges within the same topological complex that are associated to the face via the boundary relations. These can be organized as rings.
[ISO 19107:2019]

facility physical installation or physical area that may be accessed and used
EXAMPLE Elevators, restaurant areas, waiting areas, seats, toilets, shops.
NOTE 1 Such facilities may be available on-board transport means during the transport, at arrivals to and at departures from a transfer node, and at transfer nodes.
[ISO 19147:2015]

fail verdict test verdict of non-conformance
NOTE 1 Non-conformance may be with respect to either the test purpose or at least one of the conformance requirements of the relevant standard(s).
[ISO 19105:2000]

falsification test test to find errors in the implementation
NOTE 1 If errors are found, one can correctly deduce that the implementation does not conform to the International Standard; however, the absence of errors does not necessarily imply the converse. The falsification test can only demonstrate non-conformance
[ISO 19105:2000]

feature abstraction of real world phenomena
NOTE 1 A feature may occur as a type or an instance. Feature type or feature instance shall be used when only one is meant.
[ISO 19101-1:2014]

feature <UML> property of a classifier
[UML 2, ISO 19103:2015]

feature association relationship that links instances of one feature type with instances of the same or a different feature type
[ISO 19110:2016]

feature association concept concept that may be specified in detail as one or more feature association types
EXAMPLE A "supports" feature association concept describes a relationship between real world phenomena such as "highways" and "bridges" where the role of one feature is that it is supported-by the other feature (whose role is supporter-of).
[ISO 19126:2009]

feature attribute characteristic of a feature
EXAMPLE 1 A feature attribute named 'color' may have an attribute value 'green' which belongs to the data type 'text'.
EXAMPLE 2 A feature attribute named 'length' may have an attribute value '82.4' which belongs to the data type 'real'.
NOTE 1 A feature attribute has a name, a data type, and a value domain associated to it. A feature attribute for a feature instance also has an attribute value taken from the value domain.
NOTE 2 In a feature catalogue, a feature attribute may include a value domain but does not specify attribute values for feature instances.
[ISO 19101-1:2014]

feature attribute concept concept that may be specified in detail as one or more feature attribute types
EXAMPLE A "height" feature attribute concept describes length in the vertical direction as a characteristic that may be shared by real world phenomena such as "human", "tree" and "building".
[ISO 19126:2009]

feature catalogue catalogue containing definitions and descriptions of the feature types, feature attributes, and feature relationships occurring in one or more sets of geographic data, together with any feature operations that may be applied
[ISO 19101-1:2014]

feature concept concept that may be specified in detail as one or more feature types
EXAMPLE The feature concept "road" may be used to specify several different feature types, each with a different set of properties appropriate for a particular application. For a travel planning application, it might have a limited set of attributes such as name, route number, location and number of lanes, while for a maintenance application it might have an extensive set of attributes detailing the structure and composition of each of the layers of material for which it is composed.
[ISO 19126:2009]

feature concept dictionary dictionary that contains definitions of, and related descriptive information about, concepts that may be specified in detail in a feature catalogue
[ISO 19126:2009]

feature division feature succession in which a previously existing feature is replaced by two or more distinct feature instances of the same feature type
EXAMPLE An instance of the feature type "land parcel" is replaced by two instances of the same type when the parcel is legally subdivided.
[ISO 19108:2002]

feature event information about the occurrence of a located feature along a locating feature
NOTE 1 A feature event includes the linearly referenced location of the located feature along the locating feature.
NOTE 2 A feature event may be qualified by the instant in which, or period during which, the feature event occurred.
[ISO 19148:2012]

feature fusion feature succession in which two or more previously existing instances of a feature type are replaced by a single instance of the same feature type
EXAMPLE Two instances of the feature type "pasture" are replaced by a single instance when the fence between the pastures is removed.
[ISO 19108:2002]

feature identifier identifier that uniquely designates a feature instance
[ISO 19142:2010]

feature inheritance mechanism by which more specific features incorporate structure and behavior of more general features related by behavior
[ISO 19110:2016]

feature instance individual of a given feature type having specified feature attribute values
[ISO 19101-1:2014]

feature operation operation that every instance of a feature type may perform
EXAMPLE A feature operation upon a 'dam' is to raise the dam. The results of this operation are to raise the height of the 'dam' and the level of water in a 'reservoir'.
NOTE 1 Sometimes feature operations provide a basis for feature type definition.
[ISO 19110:2016]

feature operation concept concept that may be specified in detail as one or more feature operation types
EXAMPLE A "traffic flow" operation might return the number of persons or vehicles expected to move on or through some kind of transportation feature during a period of time specified as input to the operation.
[ISO 19126:2009]

feature portrayal function function that maps a geographic feature to a symbol
[ISO 19117:2012]

feature reference Uniform Resource Identifier that identifies a feature
[ISO 19143:2010]

feature substitution feature succession in which one feature instance is replaced by another feature instance of the same or different feature type
EXAMPLE An instance of feature type "building" is razed and replaced by an instance of feature type "parking lot".
[ISO 19108:2002]

feature succession replacement of one or more feature instances by other feature instances, such that the first feature instances cease to exist
[ISO 19108:2002]

feature table table where the columns represent feature attributes, and the rows represent features
[ISO 19125-2:2004]

feature type class of features having common characteristics
[ISO 19156:2011]

federated archives group of archives that has agreed to provide access to their holdings via one or more common finding aids
[ISO 14721:2012, ISO 19165-1:2018]

fiducial centre point determined on the basis of the camera fiducial marks

NOTE 1 When there are four fiducial marks, fiducial centre is the intersection of the two lines connecting the pairs of opposite fiducial marks.
[ISO 19130-1:2018]

fiducial mark index marks, typically four or eight rigidly connected with the camera body, which form images on the film negative and define the image coordinate reference system
NOTE 1 When a camera is calibrated the distances between fiducial marks are precisely measured and assigned coordinates that assist in correcting for film distortion.
[ISO 19130-1:2018]

field of regard total angular extent over which the field of view (FOV) may be positioned
NOTE 1 The field of regard is the area that is potentially able to be viewed by a system at an instant in time. It is determined by the system's FOV and the range of directions in which the system is able to point.
[Adapted from the Manual of Photogrammetry, ISO 19130-2:2014]

field of view FOV
instantaneous region seen by a sensor, provided in angular measure
NOTE 1 In the airborne case, this would be swath width for a linear array, ground footprint for an area array, and for a whiskbroom scanner it refers to the swath width.
[Manual of Photogrammetry, ISO 19130-2:2014]

file named set of records stored or processed as a unit
[ISO/IEC 2382-1:1993, ISO 19118:2011]

filter capabilities XML metadata, encoded in XML, that describes which predicates defined in this International Standard a system implements
[ISO 19143:2010]

filter expression predicate expression encoded using XML
[ISO 19143:2010]

filter expression processor component of a system that processes a filter expression
[ISO 19143:2010]

first geodetic problem direct geodetic problem
<differential geometry, geodesy> problem that given a point on a surface and the direction and distance from that point to a second point along a geodesic, determines that second point
NOTE 1 This "problem" defines a mapping from the vector space at a point (each vector given by a direction and a length) to points of the Figure of Earth that satisfy the problem for that direction and distance. For example, if we fix the distance "r" and take all directions, the resultant geometry is the circle centred at the original point of radius "r". This document will make heavy use of this mapping; see exponential map and the second geodetic problem.
[ISO 19107:2019]

first return first reflected signal that is detected by a 3D imaging system, time of flight (TOF) type, for a given sampling position and a given emitted pulse
[Adapted from STM E2544, ISO 19130-2:2014]

flattening *f*
ratio of the difference between the semi-major axis (a) and semi-minor axis (b) of an ellipsoid to the semi-major axis; $f = (a - b)/a$
NOTE 1 Sometimes inverse flattening $1/f = a/(a-b)$ is given instead; 1/f is also known as reciprocal flattening.
[ISO 19111:2019]

foliation one parameter set of geometries such that each point in the prism of the set is in one and only one trajectory and in one and only one leaf
[ISO 19141:2008]

footprint 2D extent or projection of a 3D object on a horizontal surface
[ISO 19107:2019]

format language construct that specifies the representation, in character form, of data objects in a record, file, message, storage device, or transmission channel
[ISO/IEC 2382-15:1999, ISO 19145:2013]

frame reference epoch epoch of coordinates that define a dynamic reference frame
[ISO 19111:2019]

frame sensor sensor that detects and collects all of the data for an image (frame/rectangle) at an instant of time
[ISO 19130-1:2018]

frame <LIDAR> data collected by the receiver as a result of all returns from a single emitted pulse
NOTE 1 A complete 3D data sample of the world produced by a LIDAR taken at a certain time, place, and orientation. A single LIDAR frame is also referred to as a range image.
[Adapted from NISTIR 7117, ISO 19130-2:2014]

framework relationship between the elements of the content model and the separate encoding and portrayal mechanisms
[ISO 19129:2009]

framework logical structure for classifying and organizing complex information
[ISO 27790:2009, ISO 19150-1:2012]

free function <mathematics, programming> function in an object-oriented programming language not associated to any object class
[ISO 19107:2019]

free text textual information that can be expressed in one or many languages
[ISO 19115-1:2014]

function <mathematics, programming> rule that associates each element from a domain ("source domain", or "domain" of the function) to a unique element in another domain ("target domain", "co-domain", or "range" of the function)
[ISO 19107:2019]

functional language language in which feature operations are formally specified
NOTE 1 In a functional language, feature types may be represented as abstract data types.
[ISO 19110:2016]

functional standard existing geographic information standard, in active use by an international community of data producers and data users
NOTE 1 GDF, S-57, and DIGEST are examples of functional standards.
[ISO 19101-1:2014]

fused image image produced by fusing images from multiple sources
[ISO 19163-1:2016]

gazetteer register of location instances of one or more location sub-types, containing some information regarding position
NOTE 1 The positional information need not be coordinates, but could be descriptive.
[ISO 19112:2019]

geiger mode photon counting mode for LIDAR systems, where the detector is biased and becomes sensitive to individual photons
NOTE 1 These detectors exist in the form of arrays and are bonded with electronic circuitry. The electronic circuitry produces a measurement corresponding to the time at which the current was generated; resulting in a direct time-of-flight measurement. A LIDAR that employs this detector technology typically illuminates a large scene with a single pulse. The direct time-of-flight measurements are then combined with platform location/attitude data along with pointing information to produce a three-dimensional product of the illuminated scene of interest. Additional processing is applied which removes existing noise present in the data to produce a visually exploitable dataset.
[Adapted from Albota 2002, ISO 19130-2:2014]

general concept concept which corresponds to two or more objects which form a group by reason of common properties
NOTE 1 Examples of general concepts are "planet", "tower".
[ISO 1087-1:2000, ISO 19146:2018]

generalization <UML> taxonomic relationship between a more general element and a more specific element of the same element type
NOTE 1 An instance of the more specific element can be used where the more general element is allowed. See: inheritance.
[UML 2, ISO 19103:2015]

generic concept concept in a generic relation having the narrower intension
NOTE 1 In this context, a narrower intension means fewer characteristics, resulting in a concept definition with a broader scope.
[ISO 1087-1:2000, ISO 19146:2018]

generic concept system concept system in which concepts that belong to the category of the subordinate concept are part of the extension of the superordinate concept
[ISO 19146:2018]

generic relation genus-species relation
relation between two concepts where the intension of one of the concepts includes that of the other concept and at least one additional delimiting characteristic
NOTE 1 A generic relation exists between the concepts "word" and "pronoun", "vehicle" and "car", "person" and "child".
[ISO 1087-1:2000, ISO 19146:2018]

geocentric latitude angle from the equatorial plane to the direction from the centre of an ellipsoid through a given point, northwards treated as positive
[ISO 19111:2019]

geocentric terrestrial reference system GTRS
system of geocentric space-time coordinates within the framework of General Relativity, co-rotating with the Earth and related to the Geocentric Celestial Reference System by a spatial rotation which takes into account the Earth's orientation parameters
[IAG and IUGG resolutions of 1991 and 2007, ISO 19161-1:2020]

geocoding translation of one form of location into another
NOTE 1 Geocoding usually refers to the translation of "address" or "intersection" to "direct position". Many services providers also include a "reverse geocoding" interface to their geocoder, thus extending the definition of the service as a general translator of location. Because routing services use internal location encodings not usually available to others, a geocoder is an integral part of the internals of such a service.
[ISO 19133:2005]

geodesic, geodesic line <differential geometry, geodesy> curve on a surface with a zero-length tangential curvature vector
NOTE 1 A geodesic's curvature vector is perpendicular the surface thus has the minimum curvature of any curve restricted to the surface. This is often defined as a minimal distance curve between two points, but this does not always suffice, since some points (especially on

ellipsoids and spheres) are often joined by more than one geodesic. For example, on an ellipsoid the points with $(\varphi, \lambda) = (0, 0)$ and $(0, 180)$ are joined by four separate geodesic [2 polar (the shorter) and 2 equatorial]. The exponential map is only guaranteed to be one-to-one for a small area (depending on where the centre is and how the surface is curved).
[ISO 19107:2019]

geodesic circle <differential geometry, geodesy> set of points an equal distance from a given point (on the datum)
NOTE 1 The geodesic circles centred on a pole (either one) are the lines of constant latitude. Circles in a tangent space centred on the origin (corresponding to the point of tangency) map to geodesic circles by the exponential map on the geometric reference surface centred on the point of tangency.
[ISO 19107:2019]

geodesic curvature vector tangential curvature vector
<differential geometry, geodesy> projection of the curvature vector of a curve onto the tangent plane to the surface at the point
[ISO 19107:2019]

geodetic coordinate reference system three-dimensional coordinate reference system based on a geodetic reference frame and having either a three-dimensional Cartesian or a spherical coordinate system
NOTE 1 In this document a coordinate reference system based on a geodetic reference frame and having an ellipsoidal coordinate system is geographic.
[ISO 19111:2019]

geodetic coordinate system ellipsoidal coordinate system
coordinate system in which position is specified by geodetic latitude, geodetic longitude and (in the three-dimensional case) ellipsoidal height
[ISO 19111:2019]

geodetic datum datum describing the relationship of a two- or three-dimensional coordinate system to the Earth
[ISO 19130-1:2018]

geodetic height ellipsoidal height
h
distance of a point from the reference ellipsoid along the perpendicular from the reference ellipsoid to this point, positive if upwards or outside of the reference ellipsoid
NOTE 1 Only used as part of a three-dimensional ellipsoidal coordinate system or as part of a three-dimensional Cartesian coordinate system in a three-dimensional projected coordinate reference system, but never on its own.
[ISO 19111:2019]

geodetic latitude ellipsoidal latitude
φ
angle from the equatorial plane to the perpendicular to the ellipsoid through a given point, northwards treated as positive
[ISO 19111:2019]

geodetic longitude ellipsoidal longitude
λ
angle from the prime meridian plane to the meridian plane of a given point, eastward treated as positive
[ISO 19111:2019]

geodetic reference frame reference frame or datum describing the relationship of a two- or three-dimensional coordinate system to the Earth
NOTE 1 In the data model described in this document, the UML class GeodeticReferenceFrame includes both modern terrestrial reference frames and classical geodetic datums.
[ISO 19111:2019]

geographic context awareness application or service behavior based on the recognition of user's geographic context
[ISO 19154:2014]

geographic coordinate reference system coordinate reference system that has a geodetic reference frame and an ellipsoidal coordinate system
[ISO 19111:2019]

geographic coordinates longitude, latitude and height of a ground or elevated point
NOTE 1 Geographic coordinates are related to a coordinate reference system or compound coordinate reference system. Depth equals negative height
[ISO 19130-2:2014]

geographic data data with implicit or explicit reference to a location relative to the Earth
NOTE 1 Geographic information is also used as a term for information concerning phenomena implicitly or explicitly associated with a location relative to the Earth.
[ISO 19109:2015]

geographic feature representation of real world phenomenon associated with a location relative to the Earth
[ISO 19125-2:2004]

geographic identifier spatial reference in the form of a label or code that identifies a location
EXAMPLE 'Spain' is an example of a country name, 'SW1P 3AD' is an example of a postcode.
[ISO 19112:2019]

geographic imagery imagery associated with a location relative to the Earth
[ISO 19101-2:2018]

geographic imagery scene geographic imagery whose data consists of measurements or simulated measurements of the natural world produced relative to a specified vantage point and at a specified time
NOTE 1 A geographic imagery scene is a representation of an environmental landscape; it may corre-

spond to a remotely sensed view of the natural world or to a computer-generated virtual scene simulating such a view
[ISO 19101-2:2018]

geographic information information concerning phenomena implicitly or explicitly associated with a location relative to the Earth
[ISO 19101-1:2014]

geographic information service service that transforms, manages, or presents geographic information to users
[ISO 19101-1:2014]

geographic information system information system dealing with information concerning phenomena associated with location relative to the Earth
[ISO 19101-1:2014]

geographic point location well defined geographic place described by one coordinate tuple
[ISO 19145:2013]

geographic point location representation syntactic description of a geographic point location in a well known format
[ISO 19145:2013]

geoid equipotential surface of the Earth's gravity field which is perpendicular to the direction of gravity and which best fits mean sea level either locally, regionally or globally
[ISO 19111:2019]

geolocating geopositioning an object using a Physical Sensor Model or a True Replacement Model
[ISO 19130-1:2018]

geolocation information information used to determine geographic location corresponding to image location
[ISO 19115-2:2019]

geometric aggregate collection of geometric objects that has no internal structure
NOTE 1 No assumptions about the spatial relationships between the elements can be made.
[ISO 19107:2019]

geometric boundary boundary represented by a set of geometric primitives that limits the extent of a geometric object
[ISO 19107:2019]

geometric complex set of disjoint geometric primitives where the boundary of each geometric primitive can be represented as the union of other geometric primitives of smaller dimension within the same set
NOTE 1 The geometric primitives in the set are disjoint in the sense that no direct position is interior to more than one geometric primitive. The set is closed under boundary operations, meaning that for each element in the geometric complex, there is a collection (also a geometric complex) of geometric primitives that represents the boundary of that element. Recall that the boundary of a point (the only 0D primitive object type in geometry) is empty. Thus, if the largest dimension geometric primitive is a solid (3D), the composition of the boundary operator in this definition terminates after at most three steps. It is also the case that the boundary of any object is a cycle.
[ISO 19107:2019]

geometric dimension <geometry, topology> largest number n such that each point in a set of points can be associated with a subset that has that point in its interior and is topologically isomorphic to $\mathbb{E}^n$, Euclidean n-space
NOTE 1 Curves, because they are continuous images of a portion of the real line, have geometric dimension 1. Surfaces cannot always be mapped to $\mathbb{R}^2$ in their entirety, but around each point position, a small neighbourhood can be found that resembles (under continuous functions) the interior of the unit circle in $\mathbb{R}^2$, and are therefore 2-dimensional. In this document, most surfaces (instances of Surface) are mapped to portions of $\mathbb{R}^2$ by their defining interpolation mechanisms.
[ISO 19107:2019]

geometric object spatial object representing a geometric set
NOTE 1 A geometric object consists of a geometric primitive, a collection of geometric primitives, or a geometric complex treated as a single entity. A geometric object may be the spatial representation of an object such as a feature or a significant part of a feature.
[ISO 19107:2019]

geometric primitive <geometry> geometric object representing a single, connected, homogeneous (isotropic) element of space
NOTE 1 Geometric primitives are non-decomposed objects that present information about geometric configuration. They include points, curves, surfaces, and solids. Many geometric objects behave like primitives (supporting the same interfaces defined for geometric primitives) but are actually composites composed of some number of other primitives. General collections may be aggregates and incapable of acting like a primitive (such as the lines of a complex network, which is not connected and thus incapable of being traceable as a single line). By this definition, a geometric primitive is topological open, since the boundary points are not isotropic to the interior points. Geometry is assumed to be closed. For points, the boundary is empty.
[ISO 19107:2019]

geometric realization <geometry, topology> geometric complex where the geometric primitives are in a 1-to-1 correspondence to the topological primitives of a topological complex, such that the boundary relations in the two complexes agree
NOTE 1 In such a realization, the topological primitives are considered to represent the interiors of the corresponding geometric primitives even though the primitives themselves are closed.
[ISO 19107:2019]

geometric reference surface <geometry> surface in some Euclidean space, usually $\mathbb{E}^3$, that represents an approximation to the surface of the Earth possibly restricted to a small area but often covering the entire globe
[ISO 19107:2019]

geometric set set of direct positions
NOTE 1 This set in most cases is infinite.
[ISO 19136-1:2020]

geometric set <geometry> set of points
NOTE 1 This set in most cases is infinite, except where the set consists of a list of point locations. Curves, surfaces and volumes being continuous are infinite sets of points. Some systems will define 'degenerate' curves, which are actually points.
[ISO 19107:2019]

geometry property <GML> property of a GML feature that describes some aspect of the geometry of the feature.
NOTE 1 The geometry property name is the role of the geometry in relation to the feature.
[ISO 19136-1:2020]

geometry value object object composed of a set of geometry value pairs
[ISO 19123:2005]

geometry value pair ordered pair composed of a spatial object, a temporal object or a spatiotemporal object and a record of feature attribute values
[ISO 19123:2005]

geopositioning determination of the geographic position of an object
NOTE 1 While there are many methods for geopositioning, this document is focused on geopositioning from image coordinates.
[ISO 19130-1:2018]

georectified corrected for positional displacement with respect to the surface of the Earth
[ISO 19115-2:2019]

georeferenceable associated with a geopositioning information that can be used to convert grid coordinate values to values of coordinates referenced to an external coordinate reference system related to the Earth by a datum
[ISO 19163-1:2016]

georeferencing geopositioning an object using a Correspondence Model derived from a set of points for which both ground and image coordinates are known
[ISO 19130-1:2018]

gimbal mechanical device consisting of two or more rings connected in such a way that each rotates freely around an axis that is a diameter of the next ring toward the outermost ring of the set
NOTE 1 An object mounted on a three-ring gimbal will remain horizontally suspended on a plane between the rings regardless as to the stability of the base.
[ISO 19130-1:2018]

GML application schema application schema written in XML Schema in accordance with the rules specified in this document (which is ISO 19136:2020)
[ISO 19136-1:2020]

GML document XML document with a root element that is one of the elements AbstractFeature, Dictionary or TopoComplex specified in the GML schema or any element of a substitution group of any of these elements.
[ISO 19136-1:2020]

GML profile subset of the GML schema
[ISO 19136-1:2020]

GML schema schema components in the XML namespace "http://www.opengis.net/gml/3.2" as specified in this document (which is ISO 19136:2020)
[ISO 19136-1:2020]

graphical language language whose syntax is expressed in terms of graphical symbols
[ISO 19101-1:2014]

gravity-related height H
height that is dependent on the Earth's gravity field
NOTE 1 This refers to, amongst others, orthometric height and Normal height, which are both approximations of the distance of a point above the mean sea level, but also may include Normal-orthometric heights, dynamic heights or geopotential numbers.
NOTE 2 The distance from the reference surface may follow a curved line, not necessarily straight, as it is influenced by the direction of gravity.
[ISO 19111:2019]

grazing angle <SAR> vertical angle from the local surface tangent plane to the slant range direction
[ISO 19130-1:2018]

Gregorian calendar calendar in general use; first introduced in 1582 to define a year that more closely approximated the tropical year than the Julian calendar
NOTE 1 The introduction of the Gregorian calendar included the cancellation of the accumulated inaccuracies of the Julian year. In the Gregorian calendar, a calendar year is either a common year or a leap year; each year is divided into 12 sequential months.
[ISO 8601:2000 (Adapted from), ISO 19108:2002]

grid network composed of two or more sets of curves in which the members of each set intersect the members of the other sets in an algorithmic way
NOTE 1 The curves partition a space into grid cells.
[ISO 19123:2005]

grid coordinate reference system grid CRS
coordinate reference system for the positions in a grid that uses a defined coordinate system congruent with the coordinate system described by the GridEnvelope and axisLabels of gml:GridType
NOTE 1 A grid CRS uses a defined coordinate system with the same grid point positions and origin as the

GridEnvelope, with the same axisLabels, but need not define any limits on the grid size. This coordinate system is sometimes called the internal grid coordinate system.
[ISO 19136-2:2015]

grid coordinate system coordinate system in which a position is specified relative to the intersection of curves
[ISO 19115-2:2019]

grid coordinates sequence of two or more numbers specifying a position with respect to its location on a grid
[ISO 19115-2:2019]

grid point point located at the intersection of two or more curves in a grid
[ISO 19123:2005]

gridded data data whose attribute values are associated with positions on a grid coordinate system
[ISO 19115-2:2019]

ground control point point on the earth that has an accurately known geographic position
[ISO 19115-2:2019]

ground range <SAR> magnitude of the range vector projected onto the ground
NOTE 1 Ground range of an image is represented by the distance from the nadir point of the antenna to a point in the scene. Usually measured in the horizontal plane, but can also be measured as true distance along the ground, DEM, geoid or ellipsoid surface.
[ISO 19130-1:2018]

ground reference point GRP
3D position of a reference point on the ground for a given synthetic aperture
NOTE 1 It is usually the centre point of an image (Spotlight) or an image line (Stripmap). It is usually expressed in ECEF coordinates in metres.
[ISO 19130-1:2018]

ground sampling distance linear distance between pixel centres on the ground
NOTE 1 This definition also applies for water surfaces.
[ISO 19130-1:2018]

group party any number of parties, together forming a distinct entity, with each party registered
EXAMPLE A partnership (with each partner registered as a party), or two tribes (with each tribe registered as a party).
NOTE 1 A group party may be a party member of another group party.
[ISO 19152:2012]

gyroscope device consisting of a spinning rotor mounted in a gimbal so that its axis of rotation maintains a fixed orientation
NOTE 1 The rotor spins on a fixed axis while the structure around it rotates or tilts. In airplanes, the pitch and orientation of the airplane is measured against the steady spin of the gyroscope. In space, where the four compass points are meaningless, the gyroscope's axis of rotation is used as a reference point for navigation. An inertial navigation system includes three gimbal-mounted gyroscopes, used to measure roll, pitch, and yaw.
[ISO 19130-1:2018]

heave oscillatory rise and fall of a ship due to the entire hull being lifted by the force of the sea
[IHO Hydrographic Dictionary S-32, Fifth Edition, ISO 19130-2:2014]

height distance of a point from a chosen reference surface positive upward along a line perpendicular to that surface
NOTE 1 A height below the reference surface will have a negative value.
NOTE 2 Generalisation of ellipsoidal height (h) and gravity-related height (H).
[ISO 19111:2019]

hierarchical register structured set of registers for a domain of register items, composed of a principal register and a set of subregisters
EXAMPLE ISO 6523 is associated with a hierarchical register. The principal register contains organization identifier schemes and each subregister contains a set of organization identifiers that comply with a single organization identifier scheme.
[ISO 19135-1:2015]

homograph designation having the same written form as another designation representing a different concept
EXAMPLE 1 The homographic term "die" as a noun represents different concepts in the domains of manufacturing, integrated circuits and table-top games.
EXAMPLE 2 The homographic graphical symbol (e.g. in an airport or train station) may mean "up" (e.g. an escalator) or "straight ahead" depending on the location's surroundings.
[ISO 10241-1:2011, ISO 19104:2016]

homomorphism <mathematics> relationship between two domains such that there is a structure-preserving function from one to the other
NOTE 1 A homomorphism is distinct from a isomorphism in that no inverse function is required. In an isomorphism, there are two homomorphisms that are functional inverses of one another. Continuous functions are topological homomorphisms because they preserve "topological characteristics". The mapping of topological complexes to their geometric realizations preserves the concept of boundary and is therefore a homomorphism.
[ISO 19107:2019]

homonymy relation between designations and concepts in a given language in which one designation represents two or more unrelated concepts
NOTE 1 An example of homonymy is:
bark
– 'sound made by a dog';
– 'outside covering of the stem of woody plants';
– 'sailing vessel'.

NOTE 2 The designations in the relation of homonymy are called homonyms.
[ISO 1087-1:2000, ISO 19104:2016]

homophone one of two or more words that are pronounced the same but differ in meaning, origin, and sometimes spelling
EXAMPLE night and knight.
[ISO 19104:2016]

horizontal accuracy positional accuracy of a dataset with respect to a horizontal datum
[ISO 19159-2:2016]

hydrographic swath <SONAR> strip or lane on the ground scanned by a multi-beam sounder when the survey vessel proceeds along its course
[IHO Hydrographic Dictionary S-32, Fifth Edition, ISO 19130-2:2014]

hydrophone <SONAR> component of the SONAR system which receives the sound echo and converts it to an electric signal
[ISO 19130-2:2014]

identification convention set of rules for creating identifiers
[ISO 19118:2011]

identifier linguistically independent sequence of characters capable of uniquely and permanently identifying that with which it is associated
[ISO 19135-1:2015]

identity data sufficient to identify an object over time, independent of its state
NOTE 1 An identity is usually a persistent and constant key member attribute value of the object. Since it is temporally constant and unique, it will be the same in any state associated to the object regardless of its timestamp. A moving object's identity is independent of both time and location.
[ISO 19132:2007]

image gridded coverage whose attribute values are a numerical representation of a physical parameter
NOTE 1 The physical parameters are the result of measurement by a sensor or a prediction from a model
[ISO 19115-2:2019]

image coordinate reference system coordinate reference system based on an image datum
[ISO 19111:2007]

image coordinates coordinates with respect to a Cartesian coordinate system of an image
NOTE 1 The image coordinates can be in pixels or in a measure of length (linear measure).
[ISO 19130-2:2014]

image datum engineering datum which defines the relationship of a coordinate system to an image
[ISO 19111:2007]

image distortion deviation between the actual location of an image point and the location that theoretically would result from the geometry of the imaging process without any errors
[ISO 19130-1:2018]

image formation <SAR> process by which an image is generated from collected Phase History Data in a SAR system
[ISO 19130-1:2018]

image plane plane behind an imaging lens where images of objects within the depth of field of the lens are in focus
[ISO 19130-1:2018]

image point point on the image that uniquely represents an object point
[ISO 19130-1:2018]

image-identifiable ground control point ground control point associated with a marker or other object on the ground that can be recognized in an image
NOTE 1 The ground control point may be marked in the image, or the user may be provided with an unambiguous description of the ground control point so that it can be found in the image.
[ISO 19130-1:2018]

imagery representation of phenomena as images produced by electronic and/or optical techniques
NOTE 1 In this Technical Specification, it is assumed that the phenomena have been sensed or detected by one or more devices such as radar, cameras, photometers, and infrared and multispectral scanners.
[ISO 19101-2:2018]

implementation realization of a specification
NOTE 1 In the context of the ISO geographic information standards, this includes specifications of geographic information services and datasets.
[ISO 19105:2000]

Implementation Conformance Statement ICS
statement of specification options that have been implemented
[ISO 19105:2000]

implementation coverage feature which is a subclass (specialization) of a coverage as defined in this document
NOTE 1 An implementation coverage is a concrete document in some concrete encoding, such as a GeoTIFF file.
NOTE 2 The definition of implementation coverage in this standard is functionally identical to GML coverage as defined in the OGC Coverage Implementation Schema [OGC 09-146r2].
[ISO 19123-2:2018]

Implementation eXtra Information for Testing IXIT
statement containing all of the information related to the IUT and its corresponding SUT which will enable the testing laboratory to run an appropriate test suite against that IUT
NOTE 1 IXIT typically provides the details on the organization and storage of concepts in the SUT as well as

on the means of access to and modification of the SUT.
[ISO 19105:2000]

impulse response width of the return generated by a small point reflector, which equates to the smallest distance between two point reflectors that can be distinguished as two objects
[ISO 19130-1:2018]

in situ measurement direct measurement of the measurand in its original place
[ISO 19159-1:2014]

incident angle vertical angle between the line from the detected element to the sensor and the local surface normal (tangent plane normal)
[ISO 19130-1:2018]

inconclusive verdict test verdict when neither a pass verdict nor a fail verdict apply
[ISO 19105:2000]

indirect evaluation method method of evaluating the quality of a dataset based on external knowledge
NOTE 1 Examples of external knowledge are dataset lineage, such as production method or source data.
[ISO 19157:2013]

individual instance of a class
NOTE 1 It refers to a resource belonging to the extension of the class.
[ISO 19150-2:2015]

inertial positioning system positioning system employing accelerometers, gyroscopes, and computer as integral components to determine coordinates of points or objects relative to an initial known reference point
[ISO 19116:2019]

information knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning
[ISO/IEC 2382-1:1993, ISO 19118:2011]

information package logical container composed of optional content information and optional associated preservation description information
NOTE 1 Associated with this information package is packaging information used to delimit and identify the content information and package description information used to facilitate searches for the content information.
[ISO 14721:2012, ISO 19165-1:2018]

information system information processing system, together with associated organizational resources such as human, technical, and financial resources, that provides and distributes information
[ISO/IEC 2382-1:1993, ISO 19101-1:2014]

information viewpoint viewpoint on an ODP system and its environment that focuses on the semantics of information and information processing
[ISO/IEC 10746-3:2009, ISO 19154:2014]

inheritance mechanism by which more specific classifiers incorporate structure and behavior defined by more general classifiers
NOTE 1 See generalization.
[ISO 19103:2015]

inner product <vector geometry> bilinear, symmetric function from pairs of vectors $\langle \vec{v}_1, \vec{v}_2 \rangle \longrightarrow \mathbb{R}$ to a real number such that $\langle \vec{v}, \vec{v} \rangle = \left|\left|\vec{v}\right|\right|$ and $\langle \vec{v}_1, \vec{v}_2 \rangle = \left|\left|\vec{v}_1\right|\right| \left|\left|\vec{v}_2\right|\right| \cos(\theta)$ where "θ" is the angle between $\vec{v}_1$ and $\vec{v}_2$.
NOTE 1 Inner products in differential geometry are used on the differentials that make up the local tangent spaces. In this document, this will usually be vectors tangent to a datum surface embedded a geocentric E^3 Euclidean/Cartesian space.
[ISO 19107:2019]

instance object that realizes a class
[ISO 19117:2012]

instance <UML> individual entity having its own value and possibly its own identity
NOTE 1 A classifier specifies the form and behavior of a set of instances with similar properties.
[ISO 19103:2015]

instance model representation model for storing data according to an application schema
[ISO 19118:2011]

instant 0-dimensional geometric primitive representing position in time
[ISO 19108:2002]

instantaneous field of view instantaneous region seen by a single detector element, measured in angular space
[Manual of Photogrammetry, ISO 19130-2:2014]

instantiate to represent (an abstraction) by the creation of a concrete instance or to create the ability to create an instance
NOTE 1 A class or data element definition instantiates a type if it creates the ability to create objects or data elements respectively, that can represent the concepts (instance data and /or operations) defined by that type. A class is instantiated by an object if the class defines that object's structure and function. A data schema is instantiated by a data element if the data schema defines that element's structure.
[ISO 19133:2005]

integrated positioning system positioning system incorporating two or more positioning technologies
NOTE 1 The measurements produced by each positioning technology in an integrated system may be any of position, motion, or attitude. There may be redundant measurements. When combined, a unified position, motion, or attitude is determined.
[ISO 19116:2019]

integrated side lobe ratio ISLR
ratio between the side lobe power and the main lobe

power of the impulse response of point targets in the radar imaging scene

NOTE 1 The integrated side lobe ratio (ISLR) can be obtained by integrating the power of the impulse response over suitable regions. The ISLR is expressed as

$$ISLR = 10\log_{10}\left\{\frac{P_{\text{total}} - P_{\text{main}}}{P_{\text{main}}}\right\}$$

where

P_{total} is the total power,

P_{main} is the main lobe power.

NOTE 2 The main lobe width can be taken as α times the impulse response width (IRW), centred around the peak, where α is a predefined constant, usually between 2 and 2.5.

[ISO 19159-3:2018]

integration linkage of two or more software systems by the use of a common data and method base

NOTE 1 cf. coupling. Integration and coupling are the two major mechanisms for the interoperation of systems.

[ISO 19132:2007]

intension set of characteristics which makes up the concept

[ISO 1087-1:2000, ISO 19146:2018]

intensity power per unit solid angle from a point source into a particular direction

NOTE 1 Typically for LIDAR, sufficient calibration has not been done to calculate absolute intensity, so relative intensity is usually reported. In linear mode systems, this value is typically provided as an integer, resulting from a mapping of the return's signal power to an integer value via a lookup table.

[ISO 19130-2:2014]

interface named set of operations that characterize the behavior of an entity

[ISO 19119:2016]

interface <UML> classifier that represents a declaration of a set of coherent public <UML> features and obligations

NOTE 1 An interface specifies a contract; any classifier that realizes the interface must fulfil that contract. The obligations that can be associated with an interface are in the form of various kinds of constraints (such as pre- and post-conditions) or protocol specifications, which can impose ordering restrictions on interactions through the interface.

[UML 2, ISO 19103:2015]

interferometric baseline distance between the two antenna phase centre vectors at the time when a given scatterer is imaged

[ISO 19159-3:2018]

interferometric synthetic aperture radar InSAR

technique exploiting two or more SAR images to generate maps of surface deformation or digital elevation through the differences in the phase of the waves returning to the radar

[ISO 19159-3:2018]

interior set of all direct positions that are on a geometric object but which are not on its boundary

NOTE 1 The interior of a topological object is the homomorphic image of the interior of any of its geometric realizations. This is not included as a definition because it follows from a theorem of topology.

[ISO 19136-1:2020]

internal accuracy relative accuracy

closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true

NOTE 1 Relative accuracy may also be referred to as point-to-point accuracy. The general measure of relative accuracy is an evaluation of the random errors (systematic errors and blunders removed) in determining the positional orientation (for example, distance and azimuth) of one point or feature with respect to another. In lidar, this also may specifically mean the accuracy between adjacent swaths within a lift, adjacent lifts within a project, or between adjacent projects.

[ISO 19159-2:2016]

internal coordinate reference system coordinate reference system having a datum specified with reference to the object itself

[ISO 19130-1:2018]

interoperability capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units

[ISO/IEC 2382:2009, ISO 19132:2007]

interoperate communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units

NOTE 1 cf. interoperability

[ISO 19132:2007]

interval scale scale with an arbitrary origin which can be used to describe both ordering of values and distances between values

NOTE 1 Ratios of values measured on an interval scale have no meaning.

[ISO 19108:2002]

invalidation <register> action taken to correct a substantive error in a register item

[ISO 19135-1:2015]

inverse evaluation <coverage> selection of a set of objects from the domain of a coverage based on the feature attribute values associated with the objects

[ISO 19123:2005]

irradiance electro-magnetic radiation energy per unit area per unit time
NOTE 1 The SI unit is watts per square meter (W/m^2)
[ISO 19159-1:2014]

isolated node node not related to any edge
[ISO 19107:2019]

isometry <mathematics> mapping between metric spaces that preserves the metric
NOTE 1 The formal definition is

$$f: X \to Y \ni \forall x, y \in X.\text{distance}(x, y) = \text{distance}(f(x), f(y)).$$

[ISO 19107:2019]

isomorphic <mathematics> having an isomorphism
[ISO 19107:2019]

isomorphism relationship between two domains (such as two complexes) such that there are 1-to-1, structure-preserving functions from each domain onto the other, and the composition of the two functions, in either order, is the corresponding identity function
NOTE 1 A geometric complex is isomorphic to a topological complex if their elements are in a 1-to-1, dimension- and boundary-preserving correspondence to one another.

$$[A, B \text{ isomorphic}] \Leftrightarrow [\exists f: A \to B, g: B \to A \ni f \circ g = Id_A, g \circ f = Id_B]$$

[ISO 19107:2019]

item anything that can be described and considered separately
NOTE 1 An item can be any part of a dataset, such as a feature, feature relationship, feature attribute, or combination of these.
[ISO 2859-5:2005, ISO 19157:2013]

item class set of items with common properties
NOTE 1 Class is used in this context to refer to a set of instances, not the concept abstracted from that set of instances.
[ISO 19135-1:2015]

join predicate filter expression that includes one or more clauses that constrain properties from two different entity types
NOTE 1 In this International Standard, the entity types are resource types.
[ISO 19143:2010]

join tuple set of two or more object instances that satisfy a filter that includes join predicates
NOTE 1 In this International Standard, the object instances will be feature instances.
[ISO 19142:2010]

journey movement of a person who is travelling between two locations
NOTE 1 May consist of one or more journey segments.
[ISO 19147:2015]

journey segment part of a journey defined by a start and a stop location
NOTE 1 A journey segment may be carried out by means of a trip or a subset of a trip between locations that may be transfer nodes. A journey segment may also be road use (driving, walking and cycling).
[ISO 19147:2015]

Julian date Julian day number followed by the decimal fraction of the day elapsed since the preceding noon
[ISO 19108:2002]

Julian day number number of days elapsed since Greenwich mean noon on 1 January 4713 BC, Julian proleptic calendar
[ISO 19108:2002]

junction single topological node in a network with its associated collection of turns, incoming and outgoing links
NOTE 1 Junction is an alias for node.
[ISO 19133:2005]

keystone effect distortion of a projected image caused by a tilt between the image plane and the projection plane resulting in a trapezoidal shaped projection of a rectangular image
[ISO 19159-1:2014]

knowledge cognizance which is based on reasoning
[ISO 19150-1:2012, 4.8]

knowledge base data base of knowledge about a particular subject
NOTE 1 The data base contains facts, inferences, and procedures needed for problem solution [Webster Computer]
[ISO 19101-2:2018]

land the surface of the Earth, the materials beneath, the air above and all things fixed to the soil
[UN/ECE, 2004, ISO 19152:2012]

land administration process of determining, recording and disseminating information about the relationship between people and land
NOTE 1 In many countries, land administration information is determined, recorded and disseminated under the umbrella of cadastre and land registry. Both institutions can be unified in a single (state) organization.
[ISO 19152:2012]

land cover observed (bio)physical cover on the Earth's surface
NOTE 1 Land cover is distinct from land use.
[UNFAO LCCS 2:2005, ISO 19144-2:2012]

land cover metalanguage logical general model used to describe land cover features from which more specific rules can be described to create a particular classification system
[ISO 19144-2:2012]

land use arrangements, activities and inputs people undertake in a certain land cover type to maintain it or produce change

EXAMPLE "Recreation area" is a land use term that can be applicable to different land cover types, e.g. sandy surfaces such as a beach; a built-up area such as a pleasure park; woodlands etc.

NOTE 1 The definition of land use in this way establishes a direct link between land cover and the actions of people in their environment. Multiple land uses can coexist at the same location (e.g. forestry and recreation), contrary to land cover classes that are mutually exclusive.

[UNFAO LCCS 2:2005, ISO 19144-2:2012]

language system of signs for communications, usually consisting of vocabulary and rules

NOTE 1 In this Technical Specification, language refers to natural language or special languages but not programming languages or artificial languages unless specifically identified

[ISO 5127:2001, ISO 19104:2016]

language identifier information in a terminological entry which indicates the name of a language

[ISO 1087-1:2000, ISO 19104:2016]

last return last reflected signal that is detected by a 3D imaging system, time-of-flight (TOF) type, for a given sampling position and a given emitted pulse

[Adapted from ASTM E2544, ISO 19130-2:2014]

layer basic unit of geographic information that may be requested as a map from a server

[ISO 19128:2005]

layover visual effect in SAR images of ambiguity among returns from scatterers at different heights that fall into the same range-Doppler-time bin

NOTE 1 The effect makes buildings "lay over" onto the ground toward the sensor velocity vector, akin to perspective views in projective imagery.

[ISO 19130-2:2014]

leaf <one parameter set of geometries> geometry at a particular value of the parameter

[ISO 19141:2008]

legend application of a classification in a specific area using a defined mapping scale and specific dataset

[ISO 19144-1:2009]

legend class class resultant from the application of a classification process

NOTE 1 The result of a classification process is termed legend class in this part of ISO 19144 in order to avoid confusion with the term "class" as used in UML modelling.

[ISO 19144-1:2009]

level set of spatial units, with a geometric, and/or topologic, and/or thematic coherence

EXAMPLE 1 One level of spatial units for an urban cadastre and another for spatial units for a rural cadastre.

EXAMPLE 2 One level of spatial units to define basic administrative units associated with rights and another level of spatial units to define basic administrative units associated with restrictions.

EXAMPLE 3 One level of spatial units to define basic administrative units associated with formal rights, a second level for spatial units to define basic administrative units associated with informal rights and a third level for spatial units to define basic administrative units associated with customary rights.

EXAMPLE 4 One level with point based spatial units, a second level with line based spatial units, and a third level with polygon based spatial units.

[ISO 19152:2012]

lever arm relative position vector of one sensor with respect to another in a direct georeferencing system

NOTE 1 For example, with aerial mapping cameras, there are lever arms between the inertial centre of the Inertial Measurement Unit (IMU) and the phase centre of the Global Navigation Satellite System (GNSS) antenna, each with respect to the camera perspective centre within the lens of the camera.

[ISO 19159-2:2016]

lexical language language whose syntax is expressed in terms of symbols defined as character strings

[ISO 19101-1:2014]

license permission or proof of permission granted to a system participant by a competent authority to exercise a right which would otherwise be disallowed or unlawful

[ISO 19132:2007]

lidar Light Detection and Ranging

system consisting of 1) a photon source (frequently, but not necessarily, a laser), 2) a photon detection system, 3) a timing circuit, and 4) optics for both the source and the receiver that uses emitted laser light to measure ranges to and/or properties of solid objects, gases, or particulates in the atmosphere

NOTE 1 Time of flight (TOF) LIDARs use short laser pulses and precisely record the time each laser pulse was emitted and the time each reflected return(s) is received in order to calculate the distance(s) to the scatterer(s) encountered by the emitted pulse. For topographic LIDAR, these time-of-flight measurements are then combined with precise platform location/attitude data along with pointing data to produce a three-dimensional product of the illuminated scene of interest.

[ISO 19130-2:2014]

life span period during which something exists

NOTE 1 Valid-time life span is the period during which an object exists in the modelled reality. Transaction-time life span is the period during which a database object is current in the database.

[ISO 19108:2002]

light detection and ranging LIDAR
system consisting of 1) a photon source (frequently, but not necessarily, a laser), 2) a photon detection system, 3) a timing circuit, and 4) optics for both the source and the receiver that uses emitted laser light to measure ranges to and/or properties of solid objects, gases, or particulates in the atmosphere
NOTE 1 Time of flight (TOF) LIDARs use short laser pulses and precisely record the time each laser pulse was emitted and the time each reflected return(s) is received in order to calculate the distance(s) to the scatterer(s) encountered by the emitted pulse. For topographic LIDAR, these time-of-flight measurements are then combined with precise platform location/attitude data along with pointing data to produce a three-dimensional product of the illuminated scene of interest.
[ISO 19130-2:2014]

liminal spatial unit spatial unit on the threshold between 2D and 3D representations
[ISO 19152:2012]

line string curve composed of straight-line segments
[ISO 19136-1:2020]

lineage provenance, source(s) and production process(es) used in producing a resource
[ISO 19115-1:2014]

linear coordinate system one-dimensional coordinate system in which a linear feature forms the axis
EXAMPLES Distances along a pipeline; depths down a deviated oil well bore.
[ISO 19111:2019]

linear element 1-dimensional object that serves as the axis along which linear referencing is performed
EXAMPLES Feature, such as "road"; curve geometry; directed edge topological primitive.
NOTE 1 Also known as curvilinear element.
[ISO 19148:2012]

linear mode LIDAR system in which output photocurrent is proportional to the input optical incident intensity
NOTE 1 A LIDAR system which employs this technology typically uses processing techniques to develop the time-of-flight measurements from the full waveform that is reflected from the targets in the illuminated scene of interest. These time-of-flight measurements are then combined with precise platform location/attitude data along with pointing data to produce a three-dimensional product of the illuminated scene of interest.
[ISO 19130-2:2014]

linear positioning system positioning system that measures distance from a reference point along a route
EXAMPLE An odometer used in conjunction with predefined mile or kilometre origin points along a route and provides a linear reference to a position.
[ISO 19116:2019]

linear referencing specification of a location relative to a linear element as a measurement along (and optionally offset from) that element
NOTE 1 An alternative to specifying a location as a two- or three-dimensional spatial position.
[ISO 19148:2012]

Linear Referencing Method manner in which measurements are made along (and optionally offset from) a linear element
[ISO 19148:2012]

Linear Referencing System set of Linear Referencing Methods and the policies, records and procedures for implementing them
[ISO 19148:2012]

linear segment part of a linear feature that is distinguished from the remainder of that feature by a subset of attributes, each having a single value for the entire part
NOTE 1 A linear segment is a one-dimensional object without explicit geometry.
NOTE 2 The implicit geometry of the linear segment can be derived from the geometry of the parent feature.
[ISO 19148:2012]

linearly located located using a Linear Referencing System
[ISO 19148:2012]

linearly located event occurrence along a feature of an attribute value or another feature
NOTE 1 The event location is specified using linearly referenced locations.
NOTE 2 A linearly located event may be qualified by the instant in which, or period during which, the linearly located event occurred.
NOTE 3 ISO 19108 limits events to a single instant in time and does not include the specification of a location.
[ISO 19148:2012]

linearly referenced location location whose position is specified using linear referencing
[ISO 19148:2012]

link directed topological connection between two nodes (junctions), consisting of an edge and a direction
NOTE 1 Link is an alias for directed edge.
[ISO 19133:2005]

link position position within a network on a link defined by some strictly monotonic measure associated with that link
NOTE 1 Link positions are often associated with a target feature that is not part of the network. The most common link measures used for this are the distance from start node or address. The most common use of a link position is to geolocate an "address".
[ISO 19133:2005]

linked geodata geographic data and information sources published on the Semantic Web

NOTE 1 These publicly available geographic data and information sources are published in a standardized formal model.
[ISO 19154:2014]

literal literal value
constant, explicitly specified value
NOTE 1 This contrasts with a value that is determined by resolving a chain of substitution (e.g. a variable).
[ISO 19143:2010]

literal value constant, explicitly specified value
NOTE 1 This contrasts with a value that is determined by resolving a chain of substitution (e.g. a variable)
[ISO 19143:2010]

local resource resource that is under the direct control of a system
NOTE 1 In this International Standard, the system is a web feature service and the resource is held in a data store that is directly controlled by that service.
[ISO 19142:2010]

locale definition of the subset of a user's environment that depends on language and cultural conventions
NOTE 1 In computing, a locale is a set of parameters that defines the user's language, country and any special variant preferences that the user wants to see in their user interface. Usually, a locale identifier consists of at least a language identifier and a region identifier.
[ISO/IEC IEEE 9945:2009, ISO 19160-1:2015]

localName reference to a local object directly accessible from a namespace
[ISO 19103:2015]

located feature feature that is linearly located along an associated (locating) feature
EXAMPLE A feature "bridge" may be a located feature along the feature "railway" [a locating feature].
[ISO 19148:2012]

locating feature feature that is used to identify the location of linearly located features
EXAMPLE A feature "road" may be the locating feature for a feature "pedestrian crossing" [a located feature].
[ISO 19148:2012]

location particular place or position
EXAMPLE "Madrid", "California".
NOTE 1 A location identifies a geographic place.
[ISO 19112:2019]

location based service LBS
service whose return or other property is dependent on the location of the client requesting the service or of some other thing, object or person
[ISO 19133:2005]

location dependent service LDS
service whose availability is dependent upon the location of the client
[ISO 19133:2005]

locator attribute attribute whose value is a reference to a local resource or remote resource
NOTE 1 In XML, this attribute is commonly called an href and contains a URI reference to the remote resource (see W3C XLink)
[ISO 19142:2010]

long term period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing designated community, on the information being held in an OAIS
NOTE 1 This period extends into the indefinite future.
[ISO 14721:2012, ISO 19165-1:2018]

long term preservation act of maintaining information, independently understandable by a designated community, and with evidence supporting its authenticity, over the long term
[ISO 14721:2012, ISO 19165-1:2018]

look angle vertical angle from the platform down direction to the slant range direction, usually measured at the ARP
NOTE 1 It is approximately the complement of the depression angle.
[ISO 19130-2:2014]

looks groups of signal samples in a SAR processor that splits the full synthetic aperture into several sub-apertures, each representing an independent look of the identical scene
NOTE 1 The resulting image formed by incoherent summing of these looks is characterized by reduced speckle and degraded spatial resolution.
[ISO 19163-1:2016]

loosely coupled interface message-based service interface based on a common taxonomic definition and independent of the particulars of message format or representation and of the internal implementation of the service
NOTE 1 cf. coupling
[ISO 19132:2007]

loxodrome, rhumb line <geometry, navigation> curve which crosses meridians of longitude at a constant bearing
[ISO 19107:2019]

mail recipient individual who actually receives a postal item at delivery or who first accesses the postal item if it is left for collection
NOTE 1 The mail recipient is normally the addressee, the mailee or an authorized representative of one of these two. However, this might not always be the case, e.g. if the postal item is left for collection in a location to which third parties have access; if the addressee/mailee have moved without leaving forwarding instructions or if the addressee or mailee specification was ambiguous and was, as a result, misinterpreted by the postal operator
[ISO 19160-4:2017]

mailee party designated in a postal address as having responsibility for ensuring that postal items reach their addressee
NOTE 1 Unlike the addressee, the mailee is always specified explicitly in a postal address, i.e. if a postal address does not contain a mailee, then there is no mailee.
NOTE 2 Notwithstanding Note 1 to entry, the mailee may be designated explicitly by use of a role descriptor or designated implicitly with no role descriptor.
NOTE 3 As is the case for addressee, a mailee specified in a postal address might be ambiguous.
[ISO 19160-4:2017]

mailer party who carries out one or more of the processes involved in creating, producing, finishing, inducting and paying the postage due for a postal item
[ISO 19160-4:2017]

main-road rule set of criteria used at a turn in lieu of a route instruction; default instruction used at a node
NOTE 1 This rule represents what is "most natural" to do at a node (intersection), given the entry link used. The most common version is "as straight as possible", or to exit a turn on the most obvious extension of the entry street, which is usually, but not always, the same named street that was the entry. Every node in a route is either associated with an instruction or can be navigated by the main-road rule.
[ISO 19133:2005]

management <OAIS> role played by those who set overall OAIS policy as one component in a broader policy domain, for example as part of a larger organization
[ISO 14721:2012, ISO 19165-1:2018]

maneuver, manœuvre collection of related links and turns used in a route in combination
NOTE 1 Maneuvers are used to cluster turns into convenient and legal combinations. They may be as simple as a single turn, a combination of quick turns ("jogs" in the American mid-west consisting of a turn followed immediately by a turn in the opposite direction) or very complex combinations consisting of entry, exit, and connecting roadways ("magic roundabouts" in the UK).
[ISO 19133:2005]

map portrayal of geographic information as a digital image file suitable for display on a computer screen
[ISO 19128:2005]

map projection coordinate conversion from an ellipsoidal coordinate system to a plane
[ISO 19111:2019]

matrix rectangular array of numbers
NOTE 1 A matrix is a mathematical term
[ISO 19129:2009]

maximum least upper bound, max
<mathematics> smallest element larger than or equal to all elements of a set contained in an ordered domain ($\leq$)
NOTE 1 $[\forall a \in A \Longrightarrow \max(A) \geq a] \wedge [\forall b \ni [(b \ni [\forall a \in A \Longrightarrow \mathrm{b} \geq a] \Rightarrow [\max(A) \leq b]]$
Any number is an upper bound of ∅ (empty set) as a set of numbers, because any given number is greater than any number in ∅ (an admitted vacuous statement since there is no number in ∅, but true nonetheless). This means that the max (∅) must be smaller than any number; thus "-∞".
[ISO 19107:2019]

mean sea level average level of the surface of the sea over all stages of tide and seasonal variations
NOTE 1 Mean sea level in a local context normally means mean sea level for the region calculated from observations at one or more points over a given period of time. Mean sea level in a global context differs from a global geoid by not more than 2 m.
[ISO 19111:2019]

mean sea level MSL
average height of the surface of the sea at a tide station for all stages of the tide over a 19-year period, usually determined from hourly height readings measured from a fixed predetermined reference level
[IHO Hydrographic Dictionary S-32, Fifth Edition, ISO 19130-2:2014]

measurable quantity attribute of a phenomenon, body or substance that may be distinguished qualitatively and determined quantitatively
[VIM:1993, ISO 19101-2:2018]

measurand particular quantity subject to measurement
[VIM:1993, ISO 19156:2011]

measure <GML> value described using a numeric amount with a scale or using a scalar reference system
NOTE 1 When used as a noun, measure is a synonym for physical quantity.
[ISO 19136-1:2020]

measurement set of operations having the object of determining the value of a quantity
[VIM:1993, ISO 19156:2011]

measurement accuracy accuracy of measurement, accuracy
closeness of agreement between a test result or measurement result and the true value
NOTE 1 The concept "measurement accuracy" is not a quantity and is not given a numerical quantity value. A measurement is said to be more accurate when it offers a smaller measurement error.
NOTE 2 The term "measurement accuracy" should not be used for measurement trueness and the term measurement precision should not be used for "measurement accuracy", which, however, is related to both these concepts.
NOTE 3 "Measurement accuracy" is sometimes understood as closeness of agreement between measured quantity values that are being attributed to the measurand.
[ISO 19159-1:2014]

measurement error error of measurement, error
measured quantity value minus a reference quantity value

NOTE 1 The concept of “measurement error” can be used both
a) when there is a single reference quantity value to refer to, which occurs if a calibration is made by means of a measurement standard with a measured quantity value having a negligible measurement uncertainty or if a conventional quantity value is given, in which case the measurement error is known, and
b) if a measurand is supposed to be represented by a unique true quantity value or a set of true quantity values of negligible range, in which case the measurement error is not known.
NOTE 2 Measurement error should not be confused with production error or mistake.
[ISO/IEC Guide 99:2007, ISO 19159-1:2014]

measurement precision precision
closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions
NOTE 1 Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance, or coefficient of variation under the specified conditions of measurement.
NOTE 2 The “specified conditions” can be, for example, repeatability conditions of measurement, intermediate precision conditions of measurement, or reproducibility conditions of measurement.
NOTE 3 Measurement precision is used to define measurement repeatability, intermediate measurement precision, and measurement reproducibility.
NOTE 4 Sometimes “measurement precision” is erroneously used to mean measurement accuracy.
[ISO/IEC Guide 99:2007, ISO 19159-1:2014]

medium substance or agency for storing or transmitting data
EXAMPLE Compact disc, internet, radio waves.
[ISO 19118:2011]

meridian intersection of an ellipsoid by a plane containing the shortest axis of the ellipsoid
NOTE 1 This term is often used for the pole-to-pole arc rather than the complete closed figure.
[ISO 19111:2019]

metadata information about a resource
[ISO 19115-1:2014]

metadata element discrete unit of metadata
NOTE 1 Metadata elements are unique within a metadata entity.
NOTE 2 Equivalent to an attribute in UML terminology.
[ISO 19115-1:2014]

metadata entity set of metadata elements describing the same aspect of data
NOTE 1 May contain one or more metadata entities.
NOTE 2 Equivalent to a class in UML terminology.
[ISO 19115-1:2014]

metadata section subset of metadata which consists of a collection of related metadata entities and metadata elements
NOTE 1 Equivalent to a package in UML terminology
[ISO 19115-1:2014]

metamodel model that defines the language for expressing other models
NOTE 1 A model is an instance of a metamodel, and a metamodel is an instance of a meta-metamodel.
[UML 2, ISO 19103:2015]

metaquality information describing the quality of data quality
[ISO 19157:2013]

metric operation measure
operations associated to measurements
NOTE 1 Generically, a metric is a standard of measure of any kind. The Greek word metron, meaning “measure”, giving us “metr”. Latin metricus and Greek metrikós “of, relating to measuring”.
[ISO 19107:2019]

metric traceability property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties
[Derived from VIM, ISO 19101-2:2018]

metric unit unit of measure
[ISO 19107:2019]

metrological traceability chain traceability chain
sequence of measurement standards and calibrations that is used to relate a measurement result to a reference
NOTE 1 A metrological traceability chain is defined through a calibration hierarchy.
NOTE 2 A metrological traceability chain is used to establish metrological traceability of a measurement result.
NOTE 3 A comparison between two measurement standards may be viewed as a calibration if the comparison is used to check and, if necessary, correct the quantity value and measurement uncertainty attributed to one of the measurement standards.
[ISO/IEC Guide 99:2007, ISO 19159-1:2014]

minimum greatest lower bound, min
<mathematics> largest element smaller than or equal to all elements of a set contained in an ordered domain ($\leq$)
NOTE 1
$[\forall a \in A \Longrightarrow \min(A) \leq a] \wedge [\forall b \ni [(b \ni [\forall a \in A \Longrightarrow \mathrm{b} \leq a] \Rightarrow [\min(A) \geq b]]$
Any number is a lower bound of $\varnothing$ considered as a set of numbers, because any given number is less than any number in $\varnothing$ (an admitted vacuous statement since there is no number in $\varnothing$, but true nonetheless). This means that the $\min(\varnothing)$ must be greater than any number; thus “$+\infty$”.
[ISO 19107:2019]

model abstraction of some aspects of reality
[ISO 19109:2015]

monosemy relation between designations and concepts in a given language in which one designation only relates to one concept
NOTE 1 The designations in the relation of monosemy are called monosemes.
[ISO 1087-1:2000, ISO 19146:2018]

monotonic <mathematics> never increasing or never decreasing
NOTE 1 An increasing sequence never gets smaller. A strictly increasing sequence always gets larger. A decreasing sequence never gets larger. A strictly decreasing sequence always gets smaller.

$$[x(i) \text{ is increasing}] \Leftrightarrow [i < j \Rightarrow x(i) \leq x(j)]$$
$$[x(i) \text{ is strickly increasing}] \Leftrightarrow [i < j \Rightarrow x(i) < x(j)]$$
$$[x(i) \text{ is decreasing}] \Leftrightarrow [i < j \Rightarrow x(i) \geq x(j)]$$
$$[x(i) \text{ is strickly decreasing}] \Leftrightarrow [i < j \Rightarrow x(i) > x(j)]$$
$$[x(i) \text{ is decreasing}] \vee [x(i) \text{ is increasing}] \Leftrightarrow [x(i) \text{ is monotonic}]$$

[ISO 19107:2019]

month period approximately equal in duration to the periodic time of a lunar cycle
NOTE 1 The duration of a month is an integer number of days. The number of days in a month is determined by the rules of the particular calendar.
[ISO 19108:2002]

motion change in the position of an object over time, represented by change of coordinate values with respect to a particular reference frame
EXAMPLE This may be motion of the position sensor mounted on a vehicle or other platform or motion of an object being tracked by a positioning system.
[ISO 19116:2019]

multibeam SONAR wide swath echo sounder for use in seabed mapping and surveying using the multi-beam principle
NOTE 1 Depths are measured directly below and transverse to the ship's track. The width of the swath is a function of the number of beams and their aperture.
[IHO Hydrographic Dictionary S-32, Fifth Edition, ISO 19130-2:2014]

multiple returns multiple signals returned and detected for a given emitted pulse, such as when a laser beam hitting multiple objects separated in range is split
[ISO 19130-2:2014]

multiplicity <UML> specification of the range of allowable cardinalities that a set may assume
NOTE 1 Multiplicity specifications may be given for roles within associations, parts within composites, repetitions and other purposes. Essentially a multiplicity is a (possibly infinite) subset of the non-negative integers. Contrast: cardinality.
[ISO 19103:2015]

Multipurpose Internet Mail Extensions (MIME) type media type and subtype of data in the body of a message that designates the native representation (canonical form) of such data
[ISO 19142:2010]

n-disc <topology, geometry> geometry isomorphic to the set of points X in $\mathbb{E}^n$ such that $\|X\| \leq 1$ set of all points in $\mathbb{E}^n$ less than or equal to one-unit distance from the origin
NOTE 1 The 0-disc is a point. The 1-disc is a line. The 2-disc is a circle and its interior. The 3-disc is a sphere and its interior (a ball).
[ISO 19107:2019]

n-simplex <geometry, topology> convex hull of $n + 1$ points in general position in a space of dimension at least n, or a topologically isomorphic image of such a geometry
NOTE 1 An n-simplex is n-dimensional, and is topological isomorphic to an n-disc. A 0-simplex is a point. A 1-simplex is a curve (usually a line) with two 0-simplexes (points) on its boundary; a 2-simplex is a triangular surface with three 1-simplices on its boundary. In general, an n-simplex boundary will have $n(n - 1)$-simplexes. An n-simplex and an n-disc are topologically isomorphic. The boundary of an n-simplex is isomorphic to an $(n - 1)$-sphere. In all cases, the number prefix represents the topological dimension of the object.
[ISO 19107:2019]

n-sphere <geometry, topology> geometry isomorphic to the set of points X in $\mathbb{E}^n + 1$ such that $\|X\| = 1$; set of all points in $\mathbb{E}^n + 1$ one-unit distance from the origin
NOTE 1 An n-sphere is isomorphic to the boundary of an $n + 1$ disc.
[ISO 19107:2019]

nadir point directly beneath a position
[ISO 19159-2:2016]

namespace collection of names, identified by a URI reference, that are used in XML documents as element names and attribute names
[W3C XML, ISO 19143:2010]

namespace <general> domain in which names, expressed by character strings, can be mapped to objects
NOTE 1 The names can be subject to local constraints enforced by the namespace.
[ISO 19103:2015]

namespace <RDF> common URI prefix or stem used in identifiers for a set of related resources
NOTE 1 The RDF namespace is concatenated with the localName to create the complete URI identifier for an RDF resource. Every RDF resource is identified by a URI. In contrast, an XML namespace URI scopes a local XML

component name, but there is no rule for combining these into a single identifier string.
[ISO 19150-2:2015]

narrower concept subordinate concept
concept which is either a specific concept or a partitive concept
[ISO 1087-1:2000, ISO 19146:2018]

navigation combination of routing, route traversal and tracking
[ISO 19133:2005]

navigation constraint restriction on how a link or turn may be traversed by a vehicle, such as vehicle classification, physical or temporal constraint
[ISO 19133:2005]

neighborhood <topology, metric spaces> open set of points containing a specified point in its interior
[ISO 19107:2019]

neighbourhood geometric set containing a specified direct position in its interior, and containing all direct positions within a specified distance of the specified direct position
[ISO 19132:2007]

network abstract structure consisting of a set of 0-dimensional objects called junctions, and a set of 1-dimensional objects called links that connect the junctions, each link being associated with a start (origin, source) junction and end (destination, sink) junction
NOTE 1 The network is essentially the universe of discourse for the navigation problem. Networks are a variety of 1-dimensional topological complex. In this light, junction and topological node are synonyms, as are link and directed edge.
[ISO 19133:2005]

node 0-dimensional topological primitive
NOTE 1 The boundary of a node is the empty set.
[ISO 19107:2019]

noise unwanted signal which can corrupt the measurement
NOTE 1 Noise is a random fluctuation in a signal disturbing the recognition of a carried information.
[ISO 12718:2008, ISO 19159-1:2014]

nominal value name of an object, type, or category
EXAMPLE "Deciduous needle leaf" is a nominal value that identifies a vegetation type.
NOTE 1 Many feature attributes take nominal rather than numerical values. The value domain of such an attribute is usually specified as an enumeration or a code list.
[ISO 19126:2009]

non-conformance failure to fulfil one or more specified requirements
[ISO 19105:2000]

non-verbal representation representation of a concept by means other than a descriptive statement, while revealing characteristics of this concept
NOTE 1 A non-verbal representation can be a chemical or mathematical formula, a pictographic representation or a figure, table or other kind of visual or non-visual representation revealing characteristics of the concept concerned.
[ISO 10241-1:2011, ISO 19104:2016]

normal <differential geometry, geodesy> vector perpendicular (orthogonal) to the geometric object (curve or surface) at the point
[ISO 19107:2019]

normal curvature vector <differential geometry, geodesy> projection of the curvature vector of the curve perpendicular to the tangent plane to the surface at the point
NOTE 1 The normal curvature vector of a curve is dependent only on the curve and the surface (which is a constraint on the curve). The normal curvature vector of a curve restricted to a surface is parallel to the normal vector of the surface. If the normal curvature vector is equal to the curvature vector of the curve everywhere, then the curve is a geodesic.
[ISO 19107:2019]

normal section curve <differential geometry, geodesy> plane curve segment containing the normal at one of its terminal points
NOTE 1 The usual construction begins by choosing one of the end points, the normal to the surface at that end point, and the location of the other end point and creating a plane in $\mathbb{E}^3$, which is then intersected with the datum. This curve is a normal section curve between these two points, containing the normal to the surface at the first point. On the sphere, this is the geodesic, and the two normal section curves are equal. On an ellipsoidal datum, the same is true for two points on the equator, or two points on the same meridian plane, but false in general. The reason is that in these cases the plane through the two points and the centre of the ellipsoid are planes of symmetry for the ellipsoid dividing the ellipsoid into symmetric "mirror" image halves. In general, the geodesic between the two points lies between the two normal section curves.in th
[ISO 19107:2019]

northing *N*
distance in a coordinate system, northwards (positive) or southwards (negative) from an east-west reference line
[ISO 19111:2019]

object entity with a well defined boundary and identity that encapsulates state and behavior
NOTE 1 This term was first used in this way in the general theory of object oriented programming, and later adopted for use in this same sense in UML. An object is an instance of a class. Attributes and relationships represent state. Operations, methods, and state machines represent behavior.
[UML 1, ISO 19103:2015]

object point point in the object space that is imaged by a sensor

NOTE 1 In remote sensing and aerial photogrammetry an object point is a point defined in an Earth-fixed coordinate reference system.

[ISO 19130-1:2018]

object property <OWL> semantic association between a pair of individuals

NOTE 1 Object properties have sometimes been referred to as 'abstract properties' in Description Logic.

[OWL, ISO 19150-2:2015]

objective optical element that receives light from the object and forms the first or primary image of an optical system

[ISO 19130-2:2014]

observable type data type to indicate the physical quantity as a result of an observation

[ISO 19136-1:2020]

observation act of measuring or otherwise determining the value of a property

[ISO 19156:2011]

observation procedure method, algorithm or instrument, or system of these, which may be used in making an observation

[ISO 19156:2011]

observation protocol combination of a sampling strategy and an observation procedure used in making an observation

[ISO 19156:2011]

observation result estimate of the value of a property determined through a known observation procedure

[ISO 19156:2011]

obsolete term term which is no longer in common use

[ISO 1087-1:2000]

one parameter set of geometries function f from an interval $t \in [a, b]$ such that $f(t)$ is a geometry and for each point $\mathrm{P} \in f(a)$ there is a one parameter set of points (called the trajectory of P) $\mathrm{P}(t) : [a, b] \to \mathrm{P}(t)$ such that $\mathrm{P}(t) \in f(t)$

EXAMPLE A curve C with constructive parameter t is a one parameter set of points c(t).

[ISO 19141:2008]

ontology formal representation of phenomena of a universe of discourse with an underlying vocabulary including definitions and axioms that make the intended meaning explicit and describe phenomena and their interrelationships

[ISO 19101-1:2014]

open archival information system OAIS

archive, consisting of an organization, which may be part of a larger organization, of people and systems, that has accepted the responsibility to preserve information and make it available for a designated community

NOTE 1 An OAIS Archive meets a set of responsibilities that allows to be distinguished from other uses of the term 'archive'. The term 'open' in OAIS is used to imply that this recommendation and future related recommendations and standards are developed in open forums, and it does not imply that access to the archive is unrestricted.

[ISO 14721:2012, ISO 19165-1:2018]

open set <metric, topology, geometry> containing a metric or topologically open neighborhood of each of its points

NOTE 1 In a metric space, a set X us open if each point "x" in the set is contained in some small ball which is a subset of X:

$$[X \text{ is open}] \Leftrightarrow [x \in X]$$
$$\Rightarrow [\exists \varepsilon > 0 \ni \text{distance}(x, y) < \varepsilon] \Rightarrow [y \in X].$$

A topology is a set of subsets of a space which are considered open.

[ISO 19107:2019]

open systems environment OSE

comprehensive set of interfaces, services and supporting formats, plus user aspects, for interoperability and/or portability of applications, data, or people, as specified by information technology standards and profiles

[ISO/IEC TR 10000-1:1998, ISO 19106:2004]

operating conditions parameters influencing the determination of coordinate values by a positioning system

NOTE 1 Measurements acquired in the field are affected by many instrumental and environmental factors, including meteorological conditions, computational methods and constraints, imperfect instrument construction, incomplete instrument adjustment or calibration, and, in the case of optical measuring systems, the personal bias of the observer. Solutions for positions may be affected by the geometric relationships of the observed data and/or mathematical model employed in the processing software.

[ISO 19116:2019]

operating vocabulary vocabulary that is not a reference vocabulary

[ISO 19146:2018]

operation specification of a transformation or query that an object may be called to execute

NOTE 1 An operation has a name and a list of parameters.

[ISO 19119:2016]

operation <UML> behavioral <UML> feature of a classifier that specifies the name, type, parameters, and constraints for invoking an associated behavior

[UML 2, ISO 19103:2015]

optical positioning system positioning system that determines the position of an object by means of the properties of light

EXAMPLE Total Station: Commonly used term for an integrated optical positioning system incorporating an electronic theodolite and an electronic distance-mea-

suring instrument into a single unit with an internal microprocessor for automatic computations.
[ISO 19116:2019]

ordinal era one of a set of named periods ordered in time
[ISO 19108:2002]

ordinal scale scale which provides a basis for measuring only the relative position of an object
[ISO 19108:2002]

ordinal temporal reference system temporal reference system composed of ordinal eras
[ISO 19108:2002]

orthoimage image in which by orthogonal projection to a reference surface, displacement of image points due to sensor orientation and terrain relief has been removed
NOTE 1 The amount of displacement depends on the resolution and the level of detail of the elevation information and on the software implementation.
[ISO 19101-2:2018]

package <UML> general purpose mechanism for organizing elements into groups
NOTE 1 Packages may be nested within other packages. Both model elements and diagrams may appear in a package.
[UML 2, ISO 19103:2015]

package description information intended for use by access aids
[ISO 14721:2012, ISO 19165-1:2018]

packaging information information used to bind and identify the components of an information package
EXAMPLE The ISO 9660 volume and directory information is used on a CD-ROM to provide the content of several files containing content information and preservation description information.
[ISO 14721:2012, ISO 19165-1:2018]

parameter reference epoch epoch at which the parameter values of a time-dependent coordinate transformation are valid
NOTE 1 The transformation parameter values first need to be propagated to the epoch of the coordinates before the coordinate transformation can be applied.
[ISO 19111:2019]

parameterized feature portrayal function function that maps a geographic feature to a parameterized symbol
NOTE 1 A parameterized feature portrayal function passes the relevant attribute values from the feature instance for use as input to the parameterized symbol.
[ISO 19117:2012]

parameterized symbol symbol that has dynamic parameters
NOTE 1 The dynamic parameters map to the attribute values of each feature instance being portrayed.
[ISO 19117:2012]

parametric coordinate reference system coordinate reference system based on a parametric datum
[ISO 19111:2019]

parametric coordinate system one-dimensional coordinate system where the axis units are parameter values which are not inherently spatial
[ISO 19111:2019]

parametric datum datum describing the relationship of a parametric coordinate system to an object
NOTE 1 The object is normally the Earth.
[ISO 19111:2019]

parent address address of a parent addressable object
NOTE 1 Addresses of the child addressable objects fully inherit the address components of a parent address.
[ISO 19160-1:2015]

parent addressable object addressable object that fully encloses one or more other addressable objects
EXAMPLE 1 An apartment building with many apartments within.
EXAMPLE 2 In Japan, a gaiku (block) with many jukyo bango (residence number).
EXAMPLE 3 A complex of many buildings. In Korea, a group of buildings with many dong (wings or sections of a building).
[ISO 19160-1:2015]

partition of unity <mathematics> set of real-valued functions all over the same domain whose arithmetic sum at every domain value is 1
[ISO 19107:2019]

party person or organisation that plays a role in a rights transaction
EXAMPLE An organization may be: a company, a municipality, the state, a tribe, a farmer cooperation, or a church community (with each organization represented by a delegate: a director, chief, CEO, etc.).
NOTE 1 In order to be registered as a party, not all members need to be identified and registered individually.
NOTE 2 A basic administrative unit may be a party because it may hold a right of e.g. easement.
[ISO 19152:2012]

party <postal> one or more natural and/or legal persons and/or organizations without legal personality that act(s) as a single entity for the purpose of participation in a transaction associated with a postal item
[ISO 19160-4:2017]

party member party registered and identified as a constituent of a group party
[ISO 19152:2012]

pass single instance of a remote, mobile measuring system going by a target of interest
NOTE 1 In this part of ISO 19115, the measuring system will usually be a remote sensing platform. In a navigation context, the measuring system might be a GPS satellite.
[ISO 19115-2:2019]

pass verdict test verdict of conformance
[ISO 19105:2000]

passive object object which can only react to external stimulation and cannot initiate actions on its own
NOTE 1 A passive object is usually accessed through an external interface, through which it receives requests, processes those requests and returns data as a response to that request. Since objects can implement more than one type, it is possible for a single object to pass through active and passive states. For example, a tracking service can lie dormant until a tracking request activates a period where the internals of the object initiate tracking activities based on internal triggers as specified within the request. When the tracking request is deactivated, the object may return to a passive state.
[ISO 19132:2007]

passive sensor sensor that detects and collects energy from an independent source
EXAMPLE Many optical sensors collect reflected solar energy.
[ISO 19130-1:2018]

passive SONAR type of passive sensor that only receives sound waves from external sources and does not transmit any sound waves
[ISO 19130-2:2014]

passive tracking tracking dependent on stationary sensors external to the vehicle or traveller allowing for measurements of location when the vehicle's or traveller's tracking device passes through the range of external sensors of known position
[ISO 19132:2007]

path connected property of a geometric object implying that any two points on the object can be placed on a curve that remains totally within the object
NOTE 1 There are more general geometries that are connected but not path connected, but they are not representable as collections of "digital" geometries defined in this document. No geometric objects constructible by the methods in this document on a standard digital computer can be connected but not path connected. See the notes at the term "connected"
[ISO 19107:2019]

peak side lobe ratio PSLR
ratio between the peak power of the largest side lobe and the peak power of the main lobe of the impulse response of point targets in the SAR image
NOTE 1 The peak side lobe ratio is usually expressed in decibels (dB) and computed as follows

$$PSLR = 10\log_{10}\left\{\frac{P_{\text{sidepeak}}}{P_{\text{mainpeak}}}\right\}$$

where
P_{mainpeak} is the peak power of the main lobe;
P_{sidepeak} is the peak power of the largest side lobe.
[ISO 19159-3:2018]

performance indicator internal parameters of positioning systems indicative of the level of performance achieved
NOTE 1 Performance indicators can be used as quality-control evidence of the positioning system and/or positioning solution. Internal quality control may include such factors as signal strength of received radio signals [signal-to-noise ratio (SNR)], figures indicating the dilution of precision (DOP) due to geometric constraints in radiolocation systems, and system-specific figure of merit (FOM).
[ISO 19116:2019]

performance testing measurement of the performance characteristics of an Implementation Under Test (IUT), such as its throughput, responsiveness, etc., under various conditions
NOTE 1 This is not a part of conformance testing
[ISO 19105:2000]

period one-dimensional geometric primitive representing extent in time
NOTE 1 A period is bounded by two different temporal positions.
[ISO 19108:2002]

periodic time duration of one cycle
[Adapted from ISO 31-2:1992, ISO 19108:2002]

perspective centre projection centre
point located in three dimensions through which all rays between object points and image points appear to pass geometrically
[ISO 19130-1:2018]

phase history data PHD, video phase history data
raw radar return signal information after demodulation
NOTE 1 Usually stored as a series of range lines, each containing information from a specific range bin. PHD can be thought of as a table of five columns: In-phase signal, Quadrature signal, Range, Doppler Angle, and Time.
[ISO 19130-2:2014]

physical quantity quantity used for the quantitative description of physical phenomena
NOTE 1 In GML a physical quantity is always a value described using a numeric amount with a scale or using a scalar reference system. Physical quantity is a synonym for measure when the latter is used as a noun.
[ISO 19136-1:2020]

Physical Sensor Model sensor model based on the physical configuration of a sensing system
EXAMPLE One level with point based spatial units, a second level with line based spatial units, and a third level with polygon based spatial units.
[ISO 19130-1:2018]

physiognomy general appearance of an object or terrain, without reference to its underlying or scientific characteristics
[ISO 19144-2:2012]

picture original representation of a two-dimensional hardcopy or softcopy input image in terms of the color-space coordinates (or an approximation thereof)
NOTE 1 Picture originals could be obtained from printed maps, printed pictures of a geographic imagery scene, or drawings of geographic information, etc.
[ISO 19101-2:2018]

picture portrayal representations of image data in terms of the color-space coordinates that are appropriate for, and tightly coupled to, the characteristics of a specified real or virtual output device and viewing
NOTE 1 Picture portrayals are geared for visual display whether in hardcopy or softcopy.
[ISO 19101-2:2018]

pixel smallest element of a digital image to which attributes are assigned
NOTE 1 This term originated as a contraction of "picture element"
NOTE 2 Related to the concept of a grid cell
[ISO 19101-2:2018]

pixel response non-uniformity PRNU
inhomogeneity of the response of the detectors of a detector array to a uniform activation
[ISO 19159-1:2014]

place identifiable part of any space
[ISO 19155:2012]

Place Identifier PI
reference that identifies a place
NOTE 1 The same place may be referenced by multiple Place Identifier instances. Each instance will be associated with a different reference system.
[ISO 19155:2012]

Place Identifier (PI) link relationship established between PIs and other identifiers in different encoding domains
NOTE 1 While the identifiers of these features or objects can sometimes identify a place, within the scope of this document, the identifiers of features or objects existing in other encoding domains are referred to conceptually as "other identifiers".
NOTE 2 These "other identifiers" can exist outside of the PI architecture.
[ISO 19155-2:2017]

Place Identifier (PI) linking mechanism means used to define a place identifier (PI) link
[ISO 19155-2:2017]

Place Identifier application PI application
application providing services that use Place Identifiers to end users or other applications
[ISO 19155:2012]

Place Identifier matching PI matching
matching of a Place Identifier specifying a place with another type of PI identifying the same place
NOTE 1 A source PI can be matched with multiple target Place Identifiers.
NOTE 2 PI matching can be made among coordinates, geographic identifiers and identifiers in the virtual world such as URI.
[ISO 19155:2012]

Place Identifier platform PI platform
group of service interfaces and data structures used for PI matching
[ISO 19155:2012]

planar topological complex topological complex that has a geometric realization that can be embedded in Euclidean 2 space
[ISO 19107:2019]

plane curve plane curve segment
<geometry> curve in $\mathbb{E}^3$ that is contained in a plane
NOTE 1 In a 2-dimensional coordinate system, the test for plane curve to be valid is done in the 3-dimensional space of the geometric reference surface or coordinate reference system datum.
[ISO 19107:2019]

platform structure which supports a sensor, or sensors
[ISO 19115-2:2019]

platform coordinate reference system engineering coordinate reference system fixed to the collection platform within which positions on the collection platform are defined
[ISO 19130-1:2018]

platform down direction downward normal to the platform horizontal plane
[ISO 19130-2:2014]

point 0-dimensional geometric primitive, representing a position
NOTE 1 The boundary of a point is the empty set.
[ISO 19136-1:2020]

point cloud collection of data points in 3D space
NOTE 1 The distance between points is generally non-uniform and hence all three coordinates (Cartesian or spherical) for each point must be specifically encoded.
[ISO 19130-2:2014]

point coverage coverage that has a domain composed of points
[ISO 19123:2005]

point motion operation coordinate operation that changes coordinates within one coordinate reference system due to the motion of the point
NOTE 1 The change of coordinates is from those at an initial epoch to those at another epoch.
NOTE 2 In this document the point motion is due to tectonic motion or crustal deformation.
[ISO 19111:2019]

point-spread function PSF
characteristic response of an imaging system to a high-contrast point target
[IEC 88528-11:2004, ISO 19159-1:2014]

polar coordinate system two-dimensional coordinate system in Euclidean space in which position is specified by one distance coordinate and one angular coordinate
NOTE 1 For the three-dimensional case, see spherical coordinate system.
[ISO 19111:2019]

polarimetric synthetic aperture radar SAR sensor enhanced by transmitting and receiving in different combinations of polarization
NOTE 1 By combining multiple polarization modes, it is possible to characterize the target more clearly. Quad-Pol SAR system both transmits and receives orthogonal (e.g. horizontal and vertical) polarizations, which creates four polarizations of a single imaging scene.
[ISO 19159-3:2018]

polarization restricting radiation, especially light, vibrations to a single plane
[ISO 19115-2:2019]

polarization channel imbalance bias in the estimation of the scattering matrix element ratio between coincident pixels from two coherent data channels
NOTE 1 Polarization channel imbalance includes the amplitude imbalance and phase imbalance.
[ISO 19159-3:2018]

policy set of rules related to a particular purpose
[ISO/IEC 10746-2, ISO 19101-2:2018]

polygon planar surface defined by 1 exterior boundary and 0 or more interior boundaries
[ISO 19136-1:2020]

polygon coverage coverage that has a domain composed of polygons
[ISO 19123:2005]

polymorphism characteristic of being able to assign a different meaning or usage to something in different contexts – specifically, to allow an entity such as a variable, a function, or an object to have more than one form
NOTE 1 There are several different kinds of polymorphism
[ISO 19139-1:2019]

polysemy relation between designations and concepts in a given language in which one designation represents two or more concepts sharing certain characteristics
NOTE 1 An example of polysemy is:
bridge
'structure to carry traffic over a gap';
'part of a string instrument';
'dental plate'.
NOTE 2 The designation in the relation of polysemy are called polysemes.
[ISO 1087-1:2000, ISO 19146:2018]

portrayal presentation of information to humans
[ISO 19117:2012]

portrayal catalogue collection of defined portrayals for a feature catalogue
NOTE 1 Content of a portrayal catalogue includes portrayal functions, symbols, and portrayal context (optional).
[ISO 19117:2012]

portrayal context circumstances, imposed by factors extrinsic to a geographic dataset, that affect the portrayal of that dataset
EXAMPLE Factors contributing to portrayal context can include the proposed display or map scale, the viewing conditions (day/night/dusk), and the display orientation requirements (north not necessarily at the top of the screen or page) among others.
NOTE 1 Portrayal context can influence the selection of portrayal functions and construction of symbols.
[ISO 19117:2012]

portrayal function function that maps geographic features to symbols
NOTE 1 Portrayal functions can also include parameters and other computations that are not dependent on geographic feature properties.
[ISO 19117:2012]

portrayal function set function that maps a feature catalogue to a symbol set
[ISO 19117:2012]

portrayal rule specific type of portrayal function expressed in a declarative language
NOTE 1 A declarative language is rule-based and includes decision and branching statements.
[ISO 19117:2012]

portrayal service generic interface used to portray features
[ISO 19117:2012]

position data type that describes a point or geometry potentially occupied by an object or person
NOTE 1 A direct position is a semantic subtype of position. Direct positions as described can only define a point and therefore not all positions can be represented by a direct position. That is consistent with the "is type of" relation. An 19107 geometry is also a position, just not a direct position.
[ISO 19133:2005]

positional accuracy closeness of coordinate value to the true or accepted value in a specified reference system
NOTE 1 The phrase "absolute accuracy" is sometimes used for this concept to distinguish it from relative positional accuracy. Where the true coordinate value may not be perfectly known, accuracy is normally tested by comparison to available values that can best be accepted as true.
[ISO 19116:2019]

positional reliability degree to which a positioning service provides agreed or expected absolute accuracy during a defined instant under specified conditions
[ISO 19116:2019]

positioning process computational process that determines, directly from measurements, the geodetic coordinates of points (absolute positioning), or that derives geodetic coordinates of points from previously determined geodetic coordinates (relative positioning)
[19161-1:2020]

positioning system system of instrumental and computational components for determining position
NOTE 1 Examples include inertial, integrated, linear, optical, and satellite positioning systems.
[ISO 19116:2019]

postal address address, possibly inclusive of the explicit identity of an addressee, where the addressable object is an actual or potential delivery point for a postal item
[ISO 19160-4:2017]

postal address component, component
<postal address> constituent part of a postal address
EXAMPLE Locality, postcode, thoroughfare, premises identifier
NOTE 1 A postal address component may be, but is not limited to, an element, a construct or a segment.
[ISO 19160-4:2017]

postal address construct, construct
<postal address> postal address component combining postal address elements which together form a logical portion of a postal address
[ISO 19160-4:2017]

postal address domain, domain
<postal address> an area in which a set of specific postal address types and postal address renderings is prescribed by postal operators
EXAMPLE The most typical example of a postal address domain is a country where a designated postal operator provides postal delivery services.
[ISO 19160-4:2017]

postal address element, element
<postal address> postal address component that has a well-defined conceptual meaning with significance for customer or postal processing purposes and is not itself made up of subordinate components
[ISO 19160-4:2017]

postal address element code U-code
condensed representation for a postal address element or sub-element
[ISO 19160-4:2017]

postal address rendering address rendition
<postal> process in which the rendered address is created
[ISO 19160-4:2017]

postal address segment segment
<postal address> postal address component comprising a named group of related postal address constructs and/or postal address elements with a specific defined function
[ISO 19160-4:2017]

postal address sub-element sub-element
<postal address> identifier of either a sub-division of a postal address element value or one of multiple occurrences of an element in a postal address
NOTE 1 Postal address sub-elements are used to facilitate postal address rendering, database storage and related technical needs and should not be considered as specific cases of postal address components.
[ISO 19160-4:2017]

postal address template template
<postal> specification of postal address renderings within a postal address domain
NOTE 1 Postal address template may need to include rendition instructions.
NOTE 2 A template specifies also constraints for syntactical correctness of postal addresses by indicating which elements are mandatory and which are optional.
NOTE 3 Software that interprets the rendering rules provided in template is needed to produce rendered addresses.
[ISO 19160-4:2017]

postal address type set of postal addresses composed of the same set of mandatory and optional components
NOTE 1 Postal address types may differ from country to country and from region to region within a country.
[ISO 19160-4:2017]

postal item indivisible mailable entity in respect of which a mail service contractor accepts an obligation to provide postal services
[UPU Standards Glossary, ISO 19160-4:2017]

postal operator organization licensed to provide postal services to the general public
NOTE 1 Postal administration is a special case of postal operator.
[UPU Standards Glossary, ISO 19160-4:2017]

pragmatic relation associative relation
relation between two concepts having a non-hierarchical thematic connection by virtue of experience
NOTE 1 An associative relation exists between the concepts 'education' and 'teaching', 'baking' and 'oven'.
[ISO 1087-1:2000, ISO 19146:2018]

precision measurement precision
closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions
NOTE 1 Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance, or coefficient of variation under the specified conditions of measurement.
NOTE 2 The "specified conditions" can be, for example, repeatability conditions of measurement, intermediate precision conditions of measurement, or reproducibility conditions of measurement (see ISO 5725-3).

NOTE 3 Measurement precision is used to define measurement repeatability, intermediate measurement precision, and measurement reproducibility.
NOTE 4 Sometimes “measurement precision” is erroneously used to mean measurement accuracy.
[ISO/IEC Guide 99:2007, ISO 19159-2:2016]

predicate set of computational operations applied to a data instance which evaluate to true or false
[ISO 19143:2010]

predicate expression formal syntax for describing a predicate
[ISO 19143:2010]

preferred term term rated according to the scale of the term acceptability rating as the primary term for a given concept
[ISO 1087-1:2000, ISO 19104:2016]

preservation description information PDI
information which is necessary for adequate preservation of the content information and which can be categorized as provenance, reference, fixity, context, and access rights Information
[ISO 14721:2012, ISO 19165-1:2018]

prime meridian zero meridian
meridian from which the longitudes of other meridians are quantified
[ISO 19111:2019]

principal point of autocollimation point of intersection between the image plane and the normal from the perspective centre
[ISO 19130-1:2018]

principal point of best symmetry centre of the circles of equal distortion of the lens positioned in the image plane
[ISO 19130-1:2018]

principal register register that contains a description of each of the subregisters in a hierarchical register
[ISO 19135-1:2015]

prism <one parameter set of geometries> set of points in the union of the geometries (or the union of the trajectories) of a one parameter set of geometries
NOTE 1 This is a generalization of the concept of a geometric prism that is the convex hull of two congruent polygons in 3D-space. Such polyhedrons can be viewed as a foliation of congruent polygons.
[ISO 19141:2008]

process set of interrelated or interacting activities which transforms inputs into outputs
NOTE 1 The process may be broken down further into elemental activities [sub process] as is deemed necessary to control the quality of the process.
[ISO 9000:2005, ISO 19158:2012]

producer <OAIS> role played by those persons or client systems that provide the information to be preserved
NOTE 1 This can include other OAISes or internal OAIS persons or systems.
[ISO 14721:2012, ISO 19165-1:2018]

product result of a process
[ISO 9000:2005, ISO 19158:2012]

profile set of one or more base standards or subsets of base standards, and, where applicable, the identification of chosen clauses, classes, options and parameters of those base standards, that are necessary for accomplishing a particular function
NOTE 1 A profile is derived from base standards so that by definition, conformance to a profile is conformance to the base standards from which it is derived.
[ISO/IEC TR 10000-1:1998, ISO 19106:2004]

profile <UML> definition of a limited extension to a reference metamodel with the purpose of adapting the metamodel to a specific platform or domain
[UML 2, ISO 19103:2015]

projected coordinate reference system coordinate reference system derived from a geographic coordinate reference system by applying a map projection
NOTE 1 May be two- or three-dimensional, the dimension being equal to that of the geographic coordinate reference system from which it is derived.
NOTE 2 In the three-dimensional case the horizontal coordinates (geodetic latitude and geodetic longitude coordinates) are projected to northing and easting and the ellipsoidal height is unchanged.
[ISO 19111:2019]

projection centre perspective centre
point located in three dimensions through which all rays between object points and image points appear to pass geometrically
[ISO 19130-1:2018]

property facet or attribute of an object referenced by a name
[ISO 19143:2010]

property <GML> a child element of a GML object
NOTE 1 It corresponds to feature attribute and feature association role in ISO 19109. If a GML property of a feature has an xlink:href attribute that references a feature, the property represents a feature association role.
[ISO 19136-1:2020]

property <RDF> relation between subject resources and object resources
[RDF, ISO 19150-4:2019]

property restriction <OWL> special kind of class description through the definition of constraints on values and cardinalities
[OWL, ISO 19150-2:2015]

property type characteristic of a feature type
EXAMPLE Cars (a feature type) all have a characteristic color, where “color” is a property type

NOTE 1 The value for an instance of an observable property type can be estimated through an act of observation.
NOTE 2 In chemistry-related applications, the term “determinand” or “analyte” is often used.
[ISO 19156:2011]

provenance organization or individual that created, accumulated, maintained and used records
[ISO 5127:2001, ISO 19115-1:2014]

public access open access to information sources and/or services by general public users and professional users alike
[ISO 19154:2014]

pulse repetition frequency number of times the system (e.g LIDAR) emits pulses over a given time period, usually stated in kilohertz (kHz)
[ISO 19130-2:2014]

pushbroom sensor sensor that collects a single cross-track image line at one time and constructs a larger image from a set of adjacent lines resulting from the along-track motion of the sensor
[ISO 19130-1:2018]

Pythagorean metric <Euclidean geometry> distance measure on a $\mathbb{E}^n$ coordinate space using a root-mean sum of the differences between the individual coordinate offsets
NOTE 1 $P = (p_i), Q = (q_i)$, distance $(P, Q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$. The proofs of the Pythagorean metrics all depend on the local “flatness” of the space. Cartesian coordinate space which have Pythagorean metrics are called Euclidean spaces ($\mathbb{E}^n$). In the realm of coordinate reference systems, only “Engineering Coordinate Systems” are Euclidean. Any CRS using a curved Datum are by definition non-Euclidean, and cannot “truthfully” use Pythagorean metrics except for approximation. These approximations are valid for topological statements, but not for real world measures without adjustments.
[ISO 19107:2019]

qualified cardinality <OWL> cardinality restriction that applies to literals or individuals that are connected by a data property or an object property and are instance of the qualifying range [datatype or class]
[OWL, ISO 19150-2:2015]

quality degree to which a set of inherent characteristics of an object fulfils requirements
[ISO 9000:2015, ISO 19160-3:2020]

quality assessment procedure procedure by which a customer assures that its suppliers are capable of consistently delivering the product to the required quality
NOTE 1 The assessment procedure is a second-party (customer) conformity assessment activity.
[ISO 19158:2012]

quality assessment result output of the quality assessment procedure
[ISO 19158:2012]

quality assurance part of quality management focused on providing confidence that quality requirements will be fulfilled
[ISO 9000:2005, ISO 19158:2012]

quality assurance level assurance level achieved is an outcome of the quality assessment procedure
NOTE 1 Three quality assurance levels can be achieved as part of the quality assurance framework: basic, operational and full.
[ISO 19158:2012]

quality control part of quality management focussed on fulfilling quality requirements
[ISO 9000:2005, ISO 19158:2012]

quality schema conceptual schema defining aspects of quality for geographic data
[ISO 19101-1:2014]

quantity property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed as a number and a reference
NOTE 1 A reference can be a measurement unit, a measurement procedure, a reference material, or a combination of such.
NOTE 2 Symbols for quantities are given in the ISO 80000 and IEC 80000 series Quantities and units. The symbols for quantities are written in italics. A given symbol can indicate different quantities.
NOTE 3 A quantity as defined here is a scalar. However, a vector or a tensor, the components of which are quantities, is also considered to be a quantity.
NOTE 4 The concept “quantity” may be generically divided into, e.g. “physical quantity”, “chemical quantity”, and “biological quantity”, or “base quantity” and “derived quantity”.
[ISO/IEC Guide 99:2007, ISO 19159-1:2014]

radar cross section measure of the capability of the object to scatter the transmitted radar power
NOTE 1 Radar cross section is calculated as

$$\sigma = \lim_{R\to\infty} 4\pi R^2 \frac{|E_s|^2}{|E_i|^2}$$

where
σ is the radar cross section;
E_i is the electric-field strength of the incident wave;
E_s is the electric-field strength of the scattered wave at the radar with a distance R away from the target.
NOTE 2 Radar cross section has the dimensions of area, with the unit of square metres. Usually, it is expressed in the form of a logarithm with the unit of dBsm as follows

$$\sigma_{\mathrm{dBsm}} = 10 \log_{10} \sigma$$

[ISO 19159-3:2018]

radiance at a point on a surface and in a given direction, the radiant intensity of an element of the surface, divided by

the area of the orthogonal projection of this element on a plane perpendicular to the given direction
[ISO 80000-7:2008, ISO 19101-2:2018]

radiant energy energy emitted, transferred or received as radiation
[ISO 80000-7:2008, ISO 19101-2:2018]

range set of all values a function f can take as its arguments vary over its domain
[ISO 19136:2007]

range co-domain
<mathematics> acceptable target values of a function
[ISO 19107:2019]

range <coverage> set of feature attribute values associated by a function with the elements of the domain of a coverage
[ISO 19123:2005]

range <ontology> restriction to constrain the class of objects which participate in a subject-predicate-object triple
NOTE 1 A range restriction can be thought of as a type constraint on the value of a function or range of a relation.
[ISO 19150-4:2019]

range <SAR> distance between the antenna and a distant object, synonymous with slant range
[ISO 19130-2:2014]

range bin <SAR> group of radar returns that all have the same range
[ISO 19130-1:2018]

range direction slant range direction> <SAR> direction of the range vector
[ISO 19130-1:2018]

range resolution spatial resolution in the range direction
NOTE 1 For a SAR sensor, it is usually measured in terms of the impulse response of the sensor and processing system. It is a function of the bandwidth of the pulse.
[ISO 19130-1:2018]

range vector vector from the antenna to a point in the scene
[ISO 19130-1:2018]

raster usually rectangular pattern of parallel scanning lines forming or corresponding to the display on a cathode ray tube
NOTE 1 A raster is a type of grid.
[ISO 19123:2005]

real world effect actual result of using a service, rather than merely the capability offered by a service provider
[OASIS RAF, ISO 19119:2016]

realization semantic relationship between classifiers, wherein one classifier specifies a contract that another classifier guarantees to carry out
[ISO 19139:2007]

realization <UML> specialized abstraction relationship between two sets of model elements, one representing a specification (the supplier) and the other representing an implementation of the latter (the client)
NOTE 1 Realization indicates inheritance of behavior without inheritance of structure.
[UML 2, ISO 19103:2015]

receiver hardware used to detect and record signals
NOTE 1 In LIDAR and SONAR systems, the receiver detects and records reflected pulse returns.
[ISO 19130-2:2014]

record finite, named collection of related items (objects or values)
NOTE 1 Logically, a record is a set of pairs <name, item>.
[ISO 19123:2005]

rectified grid grid for which there is an affine transformation between the grid coordinates and the coordinates of an external coordinate reference system
NOTE 1 If the coordinate reference system is related to the earth by a datum, the grid is a georectified grid.
[ISO 19123:2005]

reference ellipsoid ellipsoid
<geodesy> geometric reference surface embedded in 3D Euclidean space formed by an ellipse that is rotated about a main axis
NOTE 1 For the Earth the ellipsoid is bi-axial with rotation about the polar axis. This results in an oblate ellipsoid with the midpoint of the foci located at the nominal centre of the Earth.
[ISO 19111:2019]

reference environment geographical and cultural environment in which a concept is conceived and perceived
[ISO 19104:2016]

reference frame datum
parameter or set of parameters that realize the position of the origin, the scale, and the orientation of a coordinate system
[ISO 19111:2019]

reference information information that is used as an identifier for the content information
NOTE 1 Reference information also includes identifiers that allow outside systems to refer unambiguously to particular content information. An example of reference information is an ISBN.
[ISO 14721:2012, ISO 19165-1:2018]

reference language language specified for the development and description of concepts
EXAMPLE The reference language for the ISO/TC 211 Multi-Lingual Glossary of terms is English.
[ISO 19104:2016]

reference language subregister subregister in a hierarchical multi-lingual terminology register that contains only terminological entries in the reference language
[ISO 19104:2016]

reference model framework for understanding significant relationships among the entities of some environment,

and for the development of consistent standards or specifications supporting that environment
NOTE 1 A reference model is based on a small number of unifying concepts and can be used as a basis for education and explaining standards to a non-specialist.
[ISO 14721:2012, ISO 19101-1:2014]

reference standard measurement standard designated for the calibration of other measurement standards for quantities of a given kind in a given organization or at a given location
[ISO 19159-1:2014]

reference vocabulary vocabulary that is the basis for terminological comparisons with one or more other vocabularies
[ISO 19146:2018]

referenceable grid grid associated with a transformation that can be used to convert grid coordinate values to values of coordinates referenced to an external coordinate reference system
NOTE 1 If the coordinate reference system is related to the earth by a datum, the grid is a georeferenceable grid.
[ISO 19123:2005]

refreshment digital migration where the effect is to replace a media instance with a copy that is sufficiently exact that all archival storage hardware and software continues to run as before
[ISO 14721:2012, ISO 19165-1:2018]

register set of files containing identifiers assigned to items with descriptions of the associated items
[ISO 19135-1:2015]

register manager organization to which management of a register has been delegated by the register owner
NOTE 1 In the case of an ISO register, the register manager performs the functions of the registration authority specified in the ISO/IEC Directives.
[ISO 19135-1:2015]

register owner organization that establishes a register
[ISO 19135-1:2015]

registration assignment of a permanent, unique, and unambiguous identifier to an item
[ISO 19135-1:2015]

registry information system on which a register is maintained
[ISO 19135-1:2015]

relationship <UML> semantic connection among model elements
NOTE 1 Kinds of relationships include association, generalization, metarelationship, flow and several kinds grouped under dependency.
[UML 1, ISO 19103:2015]

relative accuracy internal accuracy
closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true
NOTE 1 Relative accuracy may also be referred to as point-to-point accuracy. The general measure of relative accuracy is an evaluation of the random errors (systematic errors and blunders removed) in determining the positional orientation (for example, distance and azimuth) of one point or feature with respect to another. In lidar, this also may specifically mean the accuracy between adjacent swaths within a lift, adjacent lifts within a project, or between adjacent projects.
[ISO 19159-2:2016]

relative position position of a point with respect to the positions of other points
NOTE 1 The spatial relationship of one point relative to another may be one-, two- or three-dimensional.
[ISO 19116:2019]

relocate <reference> update a reference to a resource that has been moved or copied to a new location
EXAMPLE A server is generating a response to a GetFeature request, it has to copy a referenced feature into the response document and the server has to "relocate" the original link contained in the referencing feature to the copy placed in the response document.
[ISO 19142:2010]

remote resource resource that is not under the direct control of a system
NOTE 1 In this International Standard, the system is a web feature service. The resource is not held in any data store that is directly controlled by that service and thus cannot be directly retrieved by the service.
[ISO 19142:2010]

remote sensing collection and interpretation of information about an object without being in physical contact with the object
[ISO 19101-2:2018]

render conversion of digital graphics data into visual form
EXAMPLE Generation of an image on a video display.
[ISO 19117:2012]

rendered address rendered postal address
postal address represented as an image in the form of a rectangular shape comprising text lines in which postal address components are separated and ordered
EXAMPLE Address on mail label, order form address, address displayed on screen
[ISO 19160-4:2017]

rendered postal address rendered address
postal address represented as an image in the form of a rectangular shape comprising text lines in which postal address components are separated and ordered
EXAMPLE Address on mail label, order form address, address displayed on screen
[ISO 19160-4:2017]

rendering parameter information item that defines the context for postal address rendering
EXAMPLE When the despatching country and delivering country of the postal item differ, it is cross-border mailing and the full name of the delivering country is required on the last line of the rendered address. Otherwise, it is domestic mailing and the name of the country is not required on the rendered address.
NOTE 1 This includes guiding of rendering of postal addresses on an external medium, such as labels, data files or screens
NOTE 2 Rendering parameters do not appear in the rendered postal address, but guide or define the rendition process.
[ISO 19160-4:2017]

rendition instruction operation which either formats, abbreviates, re-arranges or separates elements within address lines when rendering a postal address
[ISO 19160-4:2017]

repackaging digital migration in which there is an alteration in the packaging information of the AIP
[ISO 14721:2012, ISO 19165-1:2018]

replication digital migration where there is no change to the packaging information, the content information, and the PDI
NOTE 1 The bits used to represent these information objects are preserved in the transfer to the same or new media instance.
[ISO 14721:2012, ISO 19165-1:2018]

representation information information that maps a data object into more meaningful concepts
EXAMPLE 1 Representation information for a bit sequence which is a FITS file might consist of the FITS standard which defines the format plus a dictionary which defines the meaning in the file of keywords which are not part of the standard.
EXAMPLE 2 JPEG software which is used to render a JPEG file; rendering the JPEG file as bits is not very meaningful to humans but the software, which embodies an understanding of the JPEG standard, maps the bits into pixels which can then be rendered as an image for human viewing.
[ISO 14721:2012, ISO 19165-1:2018]

request invocation of an operation by a client
[ISO 19128:2005]

required relationship explicit association between either spatial units, or between basic administrative units
NOTE 1 Due to legal aspects, history of data, inaccurate geometries or missing geometries, geospatial overlay techniques may generate invalid, or no relationships between spatial units, which can be introduced by required relationships.
NOTE 2 Relationships for spatial units may be defined with ISO 19125-2 types.
[ISO 19152:2012]

resolution (of a sensor) smallest difference between indications of a sensor that can be meaningfully distinguished
NOTE 1 For imagery, resolution refers to radiometric, spectral, spatial and temporal resolutions.
[ISO 19101-2:2018]

resolution (of imagery) smallest distance between two uniformly illuminated objects that can be separately resolved in an image
[ISO 19130-2:2014]

resolution <coordinate> unit associated with the least significant digit of a coordinate
NOTE 1 Coordinate resolution may have linear or angular units depending on the characteristics of the coordinate system
[ISO 6709:2008]

resolve retrieval of a referenced resource and its insertion into a server-generated response document
NOTE 1 The insertion may be accomplished by either replacing the reference inline with a copy of the resource or by relocating the reference to point to a copy of the resource that has been placed in the response document.
[ISO 19142:2010]

resource identifiable asset or means that fulfils a requirement
EXAMPLE Dataset, dataset series, service, document, initiative, software, person or organization.
[ISO 19115-1:2014]

resource digital item controlled by a system participant
[ISO 19132:2007]

response result of an operation returned from a server to a client
[ISO 19128:2005]

response model schema defining the properties of each feature type that can appear in the response to a query operation
[ISO 19142:2010]

responsibility formal or informal obligation to do something
EXAMPLE The responsibility to clean a ditch, to keep a snow-free pavement or to remove icicles from the roof during winter, or to maintain a monument.
[ISO 19152:2012]

restriction formal or informal obligation to refrain from doing something
EXAMPLE 1 It is not allowed to build within 200 metres of a fuel station; or, a servitude or mortgage as a restriction to the ownership right.
EXAMPLE 2 Sequestration can be registered for baunit as a restriction.
[ISO 19152:2012]

retirement declaration that a register item is no longer suitable for use in the production of new data
NOTE 1 The status of the retired item changes from 'valid' to 'retired'. A retired item is kept in the register to support the interpretation of data produced before its retirement.
[ISO 19135-1:2015]

return <LIDAR> sensed signal from an emitted laser pulse which has reflected off of an illuminated scene of interest
NOTE 1 There may be multiple returns for a given emitted laser pulse.
[ISO 19130-2:2014]

rhumb line loxodrome
<geometry, navigation> curve which crosses meridians of longitude at a constant bearing
[ISO 19107:2019]

right action, activity or class of actions that a system participant may perform on or using an associated resource.
[ISO 19132:2007]

rights management control, management, allocation and tracking of the rights granted to system participants
[ISO 19132:2007]

robustness testing process of determining how well an IUT processes data which contains errors
NOTE 1 This is not a part of conformance testing
[ISO 19105:2000]

route sequence of links and / or partial links that describe a path, usually between two positions, within a network
[ISO 19133:2005]

route instruction information needed at a point along a route in a network that allows that route to be traversed
NOTE 1 To minimize the number of instructions needed to complete a route traversal, a default instruction can be assumed at junctions without specifically associated instructions. This default is called the main-road rule.
[ISO 19133:2005]

route traversal process of following a route
[ISO 19133:2005]

routing finding of optimal (minimal cost function) routes between locations in a network
[ISO 19133:2005]

row-major form <mathematics, computer science> storage mechanism for multidimensional array in linear memory, organized such that each row is stored in consecutive locations and such that the complete rows are the stored one after the other and continuing on is a similar fashion of each additional index
NOTE 1 If the indexes are (i, j) with the number of rows "r" and columns "c", then the mapping between the multidimensional locations to the linear storage locations is given by:

$$[i, j \in \mathbb{Z} \ni 1 \le i \le r; 1 \le j \le c]$$
$$\Rightarrow [(i, j) \to (i-1)c + j]$$

$$[i, j, k \in \mathbb{Z} \ni 1 \le i \le r; 1 \le j \le c; 1 \le k \le f]$$
$$\Rightarrow [(i, j, k) \to (i-1)c + (j-1)f + k]$$

NOTE 2 The matrix $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ in row major form is stored as [1 2 3 4 5 6 7 8 9].
For higher dimensions, the same pattern is applied recursively.
[ISO 19107:2019]

sampling feature feature which is involved in making observations concerning a domain feature
EXAMPLE Station, transect, section or specimen.
NOTE 1 A sampling feature is an artefact of the observational strategy, and has no significance independent of the observational campaign.
[ISO 19156:2011]

satellite ephemeris numerical representation of the trajectory of the centre of mass of an Earth orbiting artificial satellite expressed in an Earth centred terrestrial reference frame
[19161-1:2020]

satellite positioning system positioning system based upon receipt of signals broadcast from satellites
NOTE 1 In this context, satellite positioning implies the use of radio signals transmitted from "active" artificial objects orbiting the Earth and received by "passive" instruments on or near the Earth's surface to determine position, velocity, and/or attitude of an object. Examples are GPS and GLONASS.
[ISO 19116:2019]

scan set of sequential frames collected during a single full cycle of a mechanical scanner representing a cross-track excursion from one side of the field of regard to the other and back again.
[ISO 19130-2:2014]

scan mode SAR mode in which the antenna beam is steered to illuminate a swath of ground at various angles relative to flight path throughout the collection
NOTE 1 Steering the antenna also allows dwell time to be increased and provides the ability to collect strips at angles non-parallel to the flight direction and with better resolution than Stripmap mode.
[ISO 19130-1:2018]

ScanSAR mode special case of stripmap mode that uses an electronically steerable antenna to quickly change the swath being imaged during collection to collect multiple parallel swaths in one pass
[ISO 19130-1:2018]

scattering matrix matrix characterizing the scattering process at the target of interest for polarimetric SAR

NOTE 1 Scattering matrix is defined via

$$\begin{pmatrix} E_H^s \\ E_V^s \end{pmatrix} = \frac{e^{jkR}}{R} \begin{pmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{pmatrix} \begin{pmatrix} E_H^i \\ E_V^i \end{pmatrix}$$

where $\begin{pmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{pmatrix}$ is the scattering matrix;

$\begin{pmatrix} E_H^i \\ E_V^i \end{pmatrix}$ is the electronic field vector of the wave incident on the scatterer;

$\begin{pmatrix} E_H^s \\ E_V^s \end{pmatrix}$ is the electronic field vector of the scattered wave;

k is the wavenumber of the illuminating wave;

R is the distance between the target and the radar antenna.

[ISO 19159-3:2018]

scene spectral radiances of a view of the natural world as measured from a specified vantage point in space and at a specified time

NOTE 1 A scene may correspond to a remotely sensed view of the natural world or to a computer-generated virtual scene simulating such a view

[ISO 22028-1:2016, ISO 19101-2:2018]

schema formal description of a model

[ISO 19101-1:2014]

schema <XML Schema> collection of schema components within the same target namespace

EXAMPLE Schema components of W3C XML Schema are types, elements, attributes, groups, etc.

[ISO 19136-1:2020]

schema document <XML Schema> XML document containing schema component definitions and declarations

NOTE 1 The W3C XML Schema provides an XML interchange format for schema information. A single schema document provides descriptions of components associated with a single XML namespace, but several documents may describe components in the same schema, i.e. the same target namespace.

[ISO 19136-1:2020]

schema model representation model for storing schemas

EXAMPLE Representation model for a schema repository

[ISO 19118:2011]

seamless mobility continuous and intuitive access to various information sources and services regardless of protocols, networks, frequency bands, and physical environments

[ISO 19154:2014]

second geodetic problem, inverse geodetic problem <differential geometry> problem that given two points, determines the initial direction and length of a geodesic that connects them

NOTE 1 See the first geodetic problem, which reverses this operation.

[ISO 19107:2019]

segment point or polygon from a set

[ISO 19132:2007]

segment <topology, geometry> minimal subpart of a geometry, usually as part of a composite

[ISO 19107:2019]

semantic type category of objects that share some common characteristics and are thus given an identifying type name in a particular domain of discourse.

[ISO 19136-1:2020]

Semantic Web Web of data with meaning

NOTE 1 The association of meaning allows data and information to be understood and processed by automated tools as well as by people.

[ISO 19101-1:2014]

semi-major axis a

semi-diameter of the longest axis of an ellipsoid

NOTE 1 This equates to the semi-diameter of the ellipsoid measured in its equatorial plane.

[ISO 19111:2019]

semi-minor axis b

semi-diameter of the shortest axis of an ellipsoid

NOTE 1 The shortest axis coincides with the rotation axis of the ellipsoid and therefore contains both poles.

[ISO 19111:2019]

sensor element of a measuring system that is directly affected by a phenomenon, body, or substance carrying a quantity to be measured

[ISO/IEC Guide 99:2007, ISO 19130-2:2014]

sensor model description of the radiometric and geometric characteristics of a sensor

[ISO 19101-2:2018]

sensor model <geopositioning> mathematical description of the relationship between the three-dimensional object space and the 2D plane of the associated image produced by a sensor

[ISO 19130-1:2018]

sequence finite, ordered collection of related items (objects or values) that may be repeated

[ISO 19111:2019]

server a particular instance of a service

[ISO 19128:2005]

service distinct part of the functionality that is provided by an entity through interfaces

[ISO 19119:2016]

service broker application that combines or offers lower-level services for specific user needs

[ISO 19132:2007]

service chain sequence of services where, for each adjacent pair of services, occurrence of the first action is necessary for the occurrence of the second action

[ISO 19119:2016]

service metadata metadata describing the operations and geographic information available at a server
[ISO 19128:2005]

service oriented architecture SOA
software architecture consisting of coupled services
NOTE 1 The most common SOAs in use today are Web services (using SOAP, UDDI, and WSDL), CORBA and DCOM.
[ISO 19132:2007]

settlement general lowering in level of a moving vessel, relative to what its level would be were it motionless, due to the regional depression of the surface of the water in which the ship moves
NOTE 1 Settlement is not an increase in displacement.
NOTE 2 Settlement is measured as an angular tilt about the centre of gravity of the vessel.
[IHO Hydrographic Dictionary S-32, Fifth Edition, 19130-2:2014]

sexagesimal degree angle represented by a sequence of values in degrees, minutes and seconds
EXAMPLE 50.0795725 degrees is represented as 50°04'46.461" sexagesimal degrees.
NOTE 1 In the case of latitude or longitude, it may also include a character indicating hemisphere
[ISO 6709:2008]

sidescan SONAR type of SONAR that transmits sound energy from the sides of a towfish, creating a fanlike beam on either side that sweeps the seafloor, and continuously records return signals, creating a "picture" of the seafloor and any other objects
NOTE 1 Sidescan SONAR is used for imaging bottom features and targets in a wide variety of water depths.
NOTE 2 This includes synthetic aperture sidescan SONAR.
[ISO 19130-2:2014]

signature text string that specifies the name and parameters required to invoke an operation
NOTE 1 It may contain optional returned parameters. This signature is usually derived from the formal definition. This is the equivalent of the UML signature.
[ISO 19110:2016]

simple <topology, geometry> homogeneous (all points have isomorphic neighborhoods) and with a simple boundary
NOTE 1 The interior is everywhere locally isomorphic to an open disc in a Euclidean coordinate space of the appropriate dimension $D^n = \{P \mid ||P|| < 1.0\}$. The boundary is a dimension one smaller. This essentially means that the object does not intersect nor touch itself. Generally used for a curve that does not cross nor touch itself with the possible exception of boundary points. Simple closed curves are isomorphic to a circle.
[ISO 19107:2019]

simple feature feature restricted to 2D geometry with linear interpolation between vertices, having both spatial and non spatial attributes
[ISO 19125-1:2004]

simple register register containing items of a single item class
EXAMPLE A register containing terminological entries in a single specified language.
[ISO 19104:2016]

simple symbol symbol that is neither compound nor parameterized
[ISO 19117:2012]

single beam SONAR type of SONAR that produces one narrow SONAR beam directly beneath the transducer/receiver and receives a return echo from the closest object
NOTE 1 Single beam sonar is commonly called a single beam echosounder (abbr: SBES).
[ISO 19130-2:2014]

slant plane <SAR> plane that passes through the sensor velocity vector and the GRP
[ISO 19130-1:2018]

slant range <SAR> magnitude of the range vector
[ISO 19130-1:2018]

slant range direction <SAR> range direction <SAR> direction of the range vector
[ISO 19130-1:2018]

slope rate of change of elevation with respect to curve length
[ISO 19133:2005]

smile distortion centre wavelength shift of spectral channels caused by optical distortion
NOTE 1 This distortion is often simply called smile.
[ISO 19159-1:2014]

solid 3-dimensional geometric primitive, representing the continuous image of a region of Euclidean 3-space
NOTE 1 A solid is realizable locally as a three parameter set of direct positions. The boundary of a solid is the set of oriented, closed surfaces that comprise the limits of the solid.
[ISO 19123:2005]

SONAR processing system system that processes the SONAR signals to determine the geopositions of objects sensed by SONAR sensors
[ISO 19130-2:2014]

Sound Navigation And Ranging SONAR
sensor that uses sound navigation and ranging technology for sensing
[ISO 19130-2:2014]

source document providing legal and/or administrative facts on which the land administration (LA) object [right, restriction, responsibility, basic administrative unit, party, or spatial unit] is based
[ISO 19152:2012]

source document document that contains the original definition of a resource
[ISO 19150-2:2015]

spacestamp value of a spatial attribute of an object at a given time, at which time the object's state is measured and recorded
NOTE 1 See timestamp.
[ISO 19132:2007]

spatial attribute feature attribute describing the spatial representation of the feature by coordinates, mathematical functions and/or boundary topology relationships
[ISO 19117:2012]

spatial dimension, adj <topology, geometry> number of independent decisions in a coordinate system required to locate a position
NOTE 1 This definition is logically equivalent to the topological dimension of spatial projection of the CRS. It describes the space as a target for geometry. Grammatically, the term can be a noun but used to describe space, as "the spatial dimension of CRS84 is 2".
[ISO 19107:2019]

spatial dimension, noun <topology, geometry> any of the independent decisions made in a coordinate system to locate a position
NOTE 1 "A Euclidean space with a spatial dimension of 3, $\mathbb{E}^3$, usually uses axis names 'x', 'y', and 'z', Its first spatial dimension is 'x', its second is 'y' and its third is 'z'". In the context of the space, the adjective use describes the space but the singular noun can use a name for an axis separately.
[ISO 19107:2019]

spatial object object used for representing a spatial characteristic of a feature
[ISO 19107:2019]

spatial operator function or procedure that has at least one spatial parameter in its domain or range
[ISO 19107:2019]

spatial position direct position that is referenced to a 2- or 3-dimensional coordinate reference system
NOTE 1 An alternative to specifying a location as a linearly referenced location.
[ISO 19148:2012]

spatial reference description of position in the real world
NOTE 1 This may take the form of a label, code or coordinate tuple.
[ISO 19111:2019]

spatial reference system system for identifying position in the real world
[ISO 19112:2003]

spatial source source with the spatial representation of one (part of) or more spatial units
EXAMPLE A field survey sketch, an orthophoto or a satellite image with evidence of the location of boundaries (collected from the field).
[ISO 19152:2012]

spatial unit single area (or multiple areas) of land and/or water, or a single volume (or multiple volumes) of space
NOTE 1 A single area is the norm and multiple areas are the exception.
NOTE 2 Spatial units are structured in a way to support the creation and management of basic administrative units.
NOTE 3 This International Standard [ISO 19152:2012] supports either 2-dimensional (2D), 3-dimensional (3D), or mixed (2D and 3D) representations of spatial units, which may be described in text ("from this tree to that river"), or based on a single point, or represented as a set of unstructured lines, or as a surface, or as a 3D volume.
NOTE 4 In addition to spatial units represented by a single point, text, or a set of unstructured lines, a spatial unit may have an area equal to zero for administrative reasons.
[ISO 19152:2012]

spatial unit group any number of spatial units, considered as an entity
EXAMPLE Spatial units together forming an administrative zone such as a section, a canton, a municipality, a department, a province, or a country. Spatial units within a planning area.
NOTE 1 The spatial units in a spatial unit group are not necessarily continuous.
[ISO 19152:2012]

spatio-parametric coordinate reference system compound coordinate reference system in which one constituent coordinate reference system is a spatial coordinate reference system and one is a parametric coordinate reference system
NOTE 1 Normally the spatial component is "horizontal" and the parametric component is "vertical".
[ISO 19111:2019]

spatio-parametric-temporal coordinate reference system compound coordinate reference system comprised of spatial, parametric and temporal coordinate reference systems
[ISO 19111:2019]

spatio-temporal coordinate reference system compound coordinate reference system in which one constituent coordinate reference system is a spatial coordinate reference system and one is a temporal coordinate reference system
[ISO 19111:2019]

spatiotemporal domain <coverage> domain composed of spatiotemporal objects
NOTE 1 The spatiotemporal domain of a continuous coverage consists of a set of direct positions defined in relation to a collection of spatiotemporal objects.
[ISO 19123:2005]

spatiotemporal object object representing a set of direct positions in space and time
[ISO 19123:2005]

spectral resolution specific wavelength interval within the electromagnetic spectrum
EXAMPLE Band 1 of Landsat TM lies between 0.45 and 0.52 µm in the visible part of the spectrum.
[ISO 19115-2:2009]

spectral responsivity responsivity per unit wavelength interval at a given wavelength
NOTE 1 The spectral responsivity is the response of the sensor with respect to the wavelengths dependent radiance.
NOTE 2 The definition is described mathematically in IEC 60050-845. The spectral responsivity is quotient of the detector output $\mathrm{d}Y(\lambda)$ by the monochromatic detector input $\mathrm{d}X_{\mathrm{e}}(\lambda) = X_{\mathrm{e},\lambda}(\lambda) \cdot \mathrm{d}\lambda$ in the wavelength interval $\mathrm{d}\lambda$ as a function of the wavelength $\boldsymbol{s}(\lambda) = \frac{\mathrm{d}\boldsymbol{Y}(\lambda)}{\mathrm{d}\boldsymbol{X}_{\mathrm{e}}(\lambda)}$
[IEC 60050-845, ISO 19159-1:2014]

spectral width specific wavelength interval within the electromagnetic spectrum
EXAMPLE Band 1 of Landsat TM lies between 0.45 and 0.52 µm in the visible part of the spectrum.
[ISO 19115-2:2019]

spherical coordinate system three-dimensional coordinate system in Euclidean space in which position is specified by one distance coordinate and two angular coordinates
NOTE 1 Not to be confused with an ellipsoidal coordinate system based on an ellipsoid 'degenerated' into a sphere.
[ISO 19111:2019]

spotlight mode <SAR> SAR mode in which the antenna beam is steered to illuminate one area during collection
NOTE 1 Spotlight mode provides the ability to collect higher resolution SAR data over relatively smaller patches of ground surface.
[ISO 19130-1:2018]

squat effect that causes a vessel moving through water to create an area of lowered pressure under its bottom that increases the effective draught (i.e. lowers the vessel in the water)
NOTE 1 The effect is a result of Bernoulli's principle of fluid dynamics. The squat represents the increase in effective draught.
NOTE 2 For a ship underway, the change of level of the bow and stern from the still water condition in response to the elevation and depression of the water level about the hull resulting from the bow and stern wave systems.
[ISO 19130-2:2014]

squint angle <SAR> angle measured from the broadside direction vector to the range direction vector in the slant plane
[ISO 19130-1:2018]

standalone quality report free text document providing fully detailed information about data quality evaluations, results and measures used
[ISO 19157:2013]

standardization activity of establishing, with regard to actual or potential problems, provisions for common and repeated use, aimed at the achievement of the optimum degree of order in a given context
NOTE 1 In particular, the activity consists of the processes of formulating, issuing and implementing standards.
NOTE 2 Important benefits of standardization are improvement of the suitability of products, processes and services for their intended purposes, prevention of barriers to trade and facilitation of technological cooperation.
[ISO/IEC Guide 2:2004, ISO 19159-1:2014]

stare scanning mode consisting of a step stair pattern
NOTE 1 This applies to a HARLIE transceiver, based on a volume phase holographic optical element.
[ISO 19130-2:2014]

start node <topology, graph theory> node in the boundary of an edge that corresponds to the start point of that edge as a curve
[ISO 19107:2019]

start point first point of a curve
[ISO 19107:2019]

state (of an object) persistent data object reflecting the internal values of all the member attributes or measurable descriptions of a object at a given time
NOTE 1 State is usually associated to an object by its identity and to a time by a timestamp.
[ISO 19132:2007]

static conversion offline process to perform a global conversion of a large amount of data
[ISO 19145:2013]

static coordinate reference system coordinate reference system that has a static reference frame
NOTE 1 Coordinates of points on or near the crust of the Earth that are referenced to a static coordinate reference system do not change with time.
NOTE 2 Metadata for a dataset referenced to a static coordinate reference system does not require coordinate epoch information.
[ISO 19111:2019]

static datum static reference frame
reference frame in which the defining parameters exclude time evolution
[ISO 19111:2019]

static reference frame static datum
reference frame in which the defining parameters exclude time evolution
[ISO 19111:2019]

stereotype <UML> extension of an existing metaclass that enables the use of platform or domain specific terminol-

ogy or notation in place of, or in addition to, the ones used for the extended metaclass
[UML 2, ISO 19103:2015]

stop point location, e.g. a platform, at a transfer node where the transport means stop to enable the traveller to board or alight from the transport means
[ISO 19147:2015]

stray light electromagnetic radiation that has been detected but did not come directly from the IFOV
NOTE 1 Stray light may be reflected light within a telescope.
NOTE 2 This definition is valid for the optical portion of the spectrum under observation.
[ISO 19159-1:2014]

strip adjustment adjustment of observations that were made from a strip of aerial or satellite images, or lidar measurements
[ISO 19159-2:2016]

stripmap mode <SAR> SAR mode in which the antenna beam is fixed throughout the collection of an image
NOTE 1 Doppler angle in processed products is fixed for all pixels. It provides the ability to collect SAR data over strips of land over a fixed swath of ground range parallel to the direction of flight.
[ISO 19130-1:2018]

subcomplex complex all of whose elements are also in a larger complex
NOTE 1 Since the definitions of geometric complex and topological complex require only that they be closed under boundary operations, the set of any primitives of a particular dimension and below is always a subcomplex of the original, larger complex. Thus, any full planar topological complex contains an edge-node graph as a subcomplex.
[ISO 19107:2019]

subject field field of special knowledge
[ISO 1087-1:2000, ISO 19104:2016]

submission agreement agreement reached, between an OAIS and the producer, that specifies a data model, and any other arrangements needed, for the data submission session
NOTE 1 This data model identifies format/contents and the logical constructs used by the producer and how they are represented on each media delivery or in a telecommunication session.
[ISO 14721:2012, ISO 19165-1:2018]

submission information package SIP
information package that is delivered by the producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated descriptive information
[ISO 14721:2012, ISO 19165-1:2018]

submitted language language that is not the reference language
NOTE 1 Terminological entries presented in a submitted language are translated from equivalent terminological entries in the reference language.
[ISO 19104:2016]

submitted language subregister subregister in a hierarchical multi-lingual terminology register that contains only terminological entries in a single submitted language
[ISO 19104:2016]

submitting organization organization authorised by a register owner to propose changes to the content of a register
[ISO 19135-1:2015]

subordinate concept narrower concept
concept which is either a specific concept or a partitive concept
[ISO 1087-1:2000, ISO 19104:2016]

sub-process activity elements of a process
EXAMPLE In the case of photogrammetric survey, aerial triangulation can be considered a sub-process.
NOTE 1 Sub-processes can be broken down even further as is deemed necessary to control the quality of the process.
[ISO 19158:2012]

subregister part of a hierarchical register that contains items from a partition of a domain of information
[ISO 19135-1:2015]

superordinate concept broader concept
concept which is either a generic concept or a comprehensive concept
[ISO 1087-1:2000, ISO 19146:2018]

supersession <register> declaration that a register item has been retired and replaced by one or more new items
NOTE 1 The status of the replaced item changes from "valid" to "superseded".
[ISO 19135-1:2015]

supplier organization or person that provides a product
NOTE 1 The supplier can be internal or external to the customer organization.
NOTE 2 In the context of this Technical Specification, the supplier has provided a product via a process that can have some impact on quality.
[ISO 9000:2005, ISO 19158:2012]

surface 2-dimensional geometric primitive, locally representing a continuous image of a region of a plane
NOTE 1 The boundary of a surface is the set of oriented, closed curves that delineate the limits of the surface. Surfaces that are isomorphic to a sphere, or to an n-torus (a topological sphere with n "handles") have no boundary. Such surfaces are called cycles.
[ISO 19136-1:2020]

swath sensed data resulting from a single flightline of collection
[ISO 19159-2:2016]

swath <LIDAR> ground area from which return data are collected during continuous airborne LIDAR operation
NOTE 1 A typical mapping mission may consist of multiple adjacent swaths, with some overlap, and the operator will turn off the laser while the aircraft is oriented for the next swath. This term may also be referred to as a Pass.
[ISO 19130-2:2014]

sweep SONAR type of SONAR that has several single beam transducer/receivers mounted on a boom, which is then operated parallel to the water's surface and orthogonal to the vessel's direction of travel
NOTE 1 Sweep sounding is commonly called multichannel echosounding (MCES).
[ISO 19130-2:2014]

swipe set of sequential frames collected during a single half cycle of a mechanical scanner representing a cross-track excursion from one side of the field of regard to the other
[ISO 19130-2:2014]

symbol portrayal primitive that can be graphic, audible, or tactile in nature, or a combination of these
[ISO 19117:2012]

symbol component symbol that is used as a piece of a compound symbol
[ISO 19117:2012]

symbol definition technical description of a symbol
[ISO 19117:2012]

symbol reference pointer in a feature portrayal function that associates the feature type with a specific symbol
[ISO 19117:2012]

symbol set collection of symbols
NOTE 1 Symbol sets are usually designed for a community of interest to portray information of interest to the community.
[ISO 19117:2012]

synonymy relation between or among terms in a given language representing the same concept
NOTE 1 The relation of synonymy exists, for example, between deuterium and heavy hydrogen.
NOTE 2 Terms which are interchangeable in all contexts are called synonyms; if they are interchangeable only in some contexts, they are called quasi-synonyms.
[ISO 1087-1:2000, ISO 19146:2018]

Synthetic Aperture Radar SAR
imaging radar system that simulates the use of a long physical antenna by collecting multiple returns from each target as the actual antenna moves along the track
NOTE 1 The electromagnetic radiation is at microwave frequencies and is sent in pulses.
[ISO 19130-1:2018]

system of concepts concept system
set of concepts structured according to the relations among them
[ISO 1087-1:2000, ISO 19146:2018]

System Under Test SUT
computer hardware, software and communication network required to support IUT
[ISO 19105:2000]

tag <XML> markup in an XML document delimiting the content of an element
EXAMPLE <Road>.
NOTE 1 A tag with no forward slash (e.g. <Road>) is called a start-tag (also opening tag), and one with a forward slash (e.g. </Road> is called an end-tag (also closing tag).
[ISO 19136-1:2020]

tagged value <UML> attribute on a stereotype used to extend a model element
[UML 2, ISO 19103:2015]

tangent <differential geometry, calculus> direction indicating the instantaneous direction of a curve at a point
NOTE 1 The tangent is usually calculated by differentiation of a functional representation of a curve but it may be approximated by a secant (double intersection) line from the point passing through another nearby point on the curve. The closer the second point is to the first, the better the approximation of the tangent's direction.
[ISO 19107:2019]

tangent plane tangent space
collection of tangent vectors for curves passing through the point
[ISO 19107:2019]

tangent space tangent plane
collection of tangent vectors for curves passing through the point
[ISO 19107:2019]

tangent vector first derivative of a curve parameterized by arc length
NOTE 1 If $c(s) = (x(s), y(s), z(s))$ is a curve in a 3D Cartesian space ($\mathbb{E}^3$), and s is arc length along c, then the tangent is $\vec{\tau}(s) = \dot{c}(s) = (\dot{x}(s), \dot{y}(s), \dot{z}(s))$, i.e. the derivative of the coordinate values of c with respect to s. The curvature vector is $\kappa(s) = \ddot{c}(s) = (\ddot{x}(s), \ddot{y}(s), \ddot{z}(s))$.
[ISO 19107:2019]

target object or person subject to being located
NOTE 1 There is little logical difference between traveller and target except that the former is normally used for a moving object which is being tracked, and the latter is used for either an object that is not moving, or an object for which a location is needed only once. A traveller is the subject of a tracking service; a target is the subject of a locating service. Since this International Standard does not make a distinction between the protocols for these logically similar services, but does need to differentiate between the two concepts, both terms will be used as appropriate to the underlying semantics of the situation. Since all of these terms refer to entities represented by objects within the system, they can be combined with ad-

jectives defined for objects. So, an active target (a target represented by an active object) can be used to represent a moving object, since the act of motion is modifying the target's internal state and is therefore initiating actions.
[ISO 19132:2007]

technical standard standard containing the definitions of item classes requiring registration
[ISO 19135-1:2015]

technology viewpoint viewpoint on an ODP system and its environment that focuses on the choice of technology in that system
[ISO/IEC 10746-3:2009, ISO 19119:2016]

template <UML> parameterized model element
[UML 2, ISO 19103:2015]

temporal coordinate distance from the origin of the interval scale used as the basis for a temporal coordinate system
[ISO 19108:2002]

temporal coordinate reference system coordinate reference system based on a temporal datum
[ISO 19111:2019]

temporal coordinate system temporal reference system based on an interval scale on which distance is measured as a multiple of a single unit of time
[ISO 19108:2002]

temporal coordinate system <geodesy> one-dimensional coordinate system where the axis is time
[ISO 19111:2019]

temporal datum datum describing the relationship of a temporal coordinate system to an object
NOTE 1 The object is normally time on the Earth.
[ISO 19111:2019]

temporal feature association feature association characterized by a reference to time or to a temporal constraint
[ISO 19108:2002]

temporal feature operation feature operation specified as a function of time
[ISO 19108:2002]

temporal position location relative to a temporal reference system
[ISO 19108:2002]

temporal reference system reference system against which time is measured
[ISO 19108:2002]

temporal sequence ordered sequence of timestamps associated to a sequence of representations of the same object
NOTE 1 Temporal sequences are not assumed to be evenly spaced in time, nor equidistance in space. For discrete change, the default logic is to sample at temporal points of change if possible, so that the timestamp is the first temporal instance where the attributes listed have taken on that combination of values. For sake of space savings, some samples in a sequence only list those values that have changed since the immediately preceding temporal sample. For this reason, a sample should only be considered in the context of its containing sequence. For rigid motions (such as vehicle tracking), only centroid (a point value) and orientation (direction of travel) are needed for a temporal sequence describing location and spatial extent. A motion in combination with an object deformation would require more information.
[ISO 19132:2007]

term verbal designation of a general concept in a specific subject field
NOTE 1 A term may contain symbols and can have variants, e.g. different forms of spelling.
[ISO 1087-1:2000, ISO 19104:2016]

term equivalent term in another language which designates the same concept
NOTE 1 A term equivalent should be accompanied by a definition of the designated concept expressed in the same language as the term equivalent
[ISO 19104:2016]

terminological data data related to concepts or their designations
NOTE 1 The more common terminological data include entry term, definition, note, grammatical label, subject label, language identifier, country identifier and source identifier.
[ISO 1087-1:2000, ISO 19104:2016]

terminological dictionary technical dictionary
collection of terminological entries presenting information related to concepts or designations from one or more specific subject fields
[ISO 1087-1:2000, ISO 19146:2018]

terminological entry part of a terminological data collection that contains the terminological data related to one concept
[ISO 1087-1:2000, ISO 19104:2016]

terminological entry identifier unique, unambiguous, and linguistically neutral identifier assigned to a terminological entry
[ISO 19104:2016]

terminology register register of terminological entries
NOTE 1 A terminology register may be structured according to language and/or domain.
[ISO 19104:2016]

terminology repository data store or document in which terms and their associated definitions are stored or recorded
[ISO 19104:2016]

terrestrial reference frame TRF
realization of a terrestrial reference system (TRS), by specifying its origin, orientation, scale, and its time evolution
NOTE 1 The realization is achieved through a set of physical points with precisely determined coordinates in a specific coordinate system, which may include the rate of coordinate change.

NOTE 2 The realization is called static when no rates of coordinate change are defined, and kinematic when rates of coordinate change are defined without considering the underlying forces causing the motion. The realization may be called dynamic when these external forces are considered. “Dynamic” is also used colloquially to describe both the dynamic and kinematic cases without distinction.
[IERS Conventions 2010, ISO 19161-1:2020]

terrestrial reference system TRS
set of conventions defining the origin, scale, orientation and time evolution of a spatial reference system co-rotating with the Earth in its diurnal motion in space
NOTE 1 The abstract concept of a TRS is realised through a terrestrial reference frame that usually consists of a set of physical points with precisely determined coordinates and optionally their rates of change. In this document terrestrial reference frame is included within the geodetic reference frame element of the data model.
[ISO 19111:2019]

tessellation partitioning of a space into a set of conterminous subspaces having the same dimension as the space being partitioned
NOTE 1 A tessellation composed of congruent regular polygons or polyhedra is a regular tessellation. One composed of regular, but non-congruent polygons or polyhedra is a semi-regular tessellation. Otherwise the tessellation is irregular.
[ISO 19123:2005]

testing laboratory organization that carries out the conformance assessment process
[ISO 19105:2000]

thematic data gridded data whose attribute values describe characteristics of a grid coverage feature in a grid format
NOTE 1 Most gridded thematic data are derived from imagery data using geophysical/atmospheric inversion algorithms. Gridded thematic data may also be obtained from other sources such as digitization of topographic map sheets.
[ISO 19163-1:2016]

Thiessen polygon polygon that encloses one of a set of points on a plane so as to include all direct positions that are closer to that point than to any other point in the set
[ISO 19123:2005]

timestamp value of time at which an object’s state is measured and recorded
[ISO 19132:2007]

topographic LIDAR LIDAR system used to measure the topography of the ground surface
NOTE 1 Generally referring to an airborne LIDAR system
[ISO 19130-2:2014]

topological boundary <geometry, topology> boundary represented by a set of oriented topological primitives of smaller topological dimension that limits the extent of a topological object, or by the geometric realization of such a set
NOTE 1 The boundary of a topological complex is the boundary of the geometric realization of the topological complex.
[ISO 19107:2019]

topological complex collection of topological primitives that is closed under the boundary operations
NOTE 1 Closed under the boundary operations means that if a topological primitive is in the topological complex, then its boundary objects are also in the topological complex.
[ISO 19107:2019]

topological dimension minimum number of free variables needed to distinguish nearby direct positions within a geometric object from one another
[ISO 19123:2005]

topological expression collection of oriented topological primitives which is operated upon like a multivariate polynomial
NOTE 1 Topological expressions are used for many calculations in computational topology.
[ISO 19107:2019]

topological object spatial object representing spatial characteristics that are invariant under continuous transformations
NOTE 1 A topological object is a topological primitive, a collection of topological primitives, or a topological complex.
[ISO 19107:2019]

topological primitive <geometry, topology> topological object that represents a single, homogeneous, non-decomposable element
NOTE 1 A topological primitive corresponds to the interior of a geometric primitive of the same dimension in a geometric realization.
[ISO 19107:2019]

topological solid 3-dimensional topological primitive
NOTE 1 The boundary of a topological solid consists of a set of directed faces.
[ISO 19107:2019]

traceability chain metrological traceability chain
sequence of measurement standards and calibrations that is used to relate a measurement result to a reference
NOTE 1 A metrological traceability chain is defined through a calibration hierarchy.
NOTE 2 A metrological traceability chain is used to establish metrological traceability of a measurement result.
NOTE 3 A comparison between two measurement standards may be viewed as a calibration if the comparison is used to check and, if necessary, correct the quantity value and measurement uncertainty attributed to one of

the measurement standards.
[ISO/IEC Guide 99:2007, ISO 19159-1:2014]

tracking monitoring and reporting the location of a vehicle
[ISO 19133:2005]

tracking device device (tag) carried by a vehicle to allow it to determine its location or to be sensed by external objects of known location
NOTE 1 The most common tracking devices are cell phones, GNSS chips, RFID (Radio Frequency ID) tags, or printed tags which are scannable by optical sensors such as “bar codes”.
NOTE 2 The common usage of “vehicle” means a “form of conveyance” or, more simply “thing that conveys (carries) something else”. Thus, a tracked object that carries a tracking device to allow it to be tracked is, by definition, a conveyance or vehicle for that device. Thus, a cell phone that carries a GNSS device is the vehicle for that device, and the traveller carrying the cell phone, allowing him to be tracked, is the vehicle for the phone and all of its internal electronics.
[ISO 19132:2007]

trajectory path of a moving point described by a one parameter set of points
[ISO 19141:2008]

transaction time time when a fact is current in a database and may be retrieved
[ISO 19108:2002]

transducer device that converts one type of energy to another
[ISO 19130-2:2014]

transfer person’s activity to switch between transport modes, transport networks or transport means
[ISO 19147:2015]

transfer link link that connects transfer nodes or stop points within a transfer node
NOTE 1 A transfer link enables travellers to move between the different transfer nodes and stop points within a transfer node.
[ISO 19147:2015]

transfer node location that facilitates transfers between transport modes, transport networks and/or transport means
NOTE 1 A transfer node may contain other transfer nodes and may be related to one or more transport modes and transport networks. It may also contain stop points and facilities for the users of the transfer node. A transfer node may host services that are provided to the users of the transfer node, e.g. information services, ticket sales, etc.
NOTE 2 A transfer node may be a part of a hierarchy of transfer nodes. Thus, a transfer node may be related to many transport modes and transport networks. However, only transfer nodes that are related to just one transport network will have stop points.
NOTE 3 The stop points related to different transfer nodes, which may serve different transport modes and networks, may, in real life, have the same physical locations. A tram and a bus may for example share the same platform, but conceptually they may belong to different transfer nodes.
[ISO 19147:2015]

transfer protocol common set of rules for defining interactions between distributed systems
[ISO 19118:2011]

transfer unit collection of data for the purpose of a data transfer
NOTE 1 A transfer unit does not have to be identifiable like a dataset.
[ISO 19118:2011]

transformation <OAIS> digital migration in which there is an alteration to the content information or PDI of an archival information package
EXAMPLE Changing ASCII codes to UNICODE in a text document being preserved is a transformation.
[ISO 14721:2012, ISO 19165-1:2018]

transformation reference epoch epoch at which the parameter values of a time-specific coordinate transformation are valid
NOTE 1 Coordinates first need to be propagated to this epoch before the coordinate transformation is applied. This is in contrast to a parameter reference epoch where the transformation parameter values first need to be propagated to the epoch of the coordinates before the coordinate transformation is applied.
[ISO 19111:2019]

transmitter component of SONAR that converts an electrical impulse into a sound wave and sends the wave into the water
NOTE 1 Transmitter is also called projector in multibeam echosounding.
[ISO 19130-2:2014]

transport means any type of vehicle, associated with any transport mode, that is used for the transport of persons or goods
[ISO 19147:2015]

transport mode means that travellers can choose for transport
NOTE 1 In this International Standard, the term ”transport mode” is frequently shortened to “mode” for convenience. Transport mode is usually covering groups of vehicles, e.g. transport means used for road transport, rail transport, sea transport, air transport and cable transport. A mode for inland water-borne transport is also added.
[ISO 19134:2007]

transport network physical network infrastructure for mobility of transport means, containing infrastructure and equipment that facilitate traffic management

NOTE 1 A transport network is associated to a specific transport mode and specific types of transport means. The road network, consisting of streets in a city may for example support several transport networks such as transport networks for pedestrians, bicycles and use private cars. The network served by a bus service is also a transport network.
[ISO 19147:2015]

transport service service that is offered to a person with a transport demand
NOTE 1 This may be services supporting mobility (e.g. public transport services) or services such as luggage handling, assistance, etc.
[ISO 19147:2015]

transportation mode means that travellers can choose for transportation
[ISO 19134:2007]

traversable condition of a link or turn that allows or restricts all traffic's traversal, as opposed to a more detailed navigation constraint
NOTE 1 Traversability is usually a function of physical, cultural, or legal conditions. If traversable is false, then the object cannot be navigated. This effectively removes a link from the usable network. In the case of a node, it effectively removes the node and all associated links from the useable network. In the case of a turn, it simply removes it from any viable route. Non-traversable entities are not included in maneuvers or routes.
[ISO 19133:2005]

traversal <XML> using or following an XLink link for any purpose
[W3C XLink:2001, ISO 19142:2010]

traversal order sequence in which the cells of a grid are enumerated
[ISO 19129:2009]

triangulated irregular network tessellation composed of triangles
[ISO 19123:2005]

trip instance of a transport service supporting mobility, for example, a specific flight, a specific ferry departure or a specific taxi tour
NOTE 1 May follow a pre-defined trip pattern or the trip may be ad hoc according to specific mobility demands. Depending on the nature of the trip, it may start and stop at several transfer nodes.
NOTE 2 A trip may serve many travellers, and the traveller may make use of different parts of the trip (e.g. different legs between different bus stops), and these parts may be whole journeys or journey segments.
[ISO 19147:2015]

trip pattern pre-defined path defined by means of two or more transfer nodes and the links and waypoints in between
NOTE 1 Defines a path to be followed by a transport service; also called route.
[ISO 19147:2015]

True Replacement Model model using functions whose coefficients are based on a Physical Sensor Model
[ISO 19130-1:2018]

tuple ordered list of values
NOTE 1 The number of values in a tuple is immutable
[ISO 19136-1:2020]

turn part of a route or network consisting of a junction location and an entry and exit link for that junction
[ISO 19133:2005]

type <UML> stereotyped class that specifies a domain of objects together with the operations applicable to the objects, without defining the physical implementation of those objects
NOTE 1 A type may have attributes and associations.
[UML 1, ISO 19103:2015]

type coercion <programming> conversion of one type of value to a value of a different type with similar content
NOTE 1 Point and DirectPosition are informationally identical (related to each other in a one to one fashion) in this context. Given a DirectPosition, a Point can be constructed. Given a Point, a DirectPosition can be derived from of its coordinates. If coercion is supported, a Point may be used where a DirectPosition is requested, and vice versa. Most programming languages use coercion, but others use "cast" operators (a type of constructor) requiring the programmer to initiate the coercion. "Strong substantiality" is related in that a subtype instance can always be coerced to any of its supertypes.
[ISO 19107:2019]

ubiquitous geographic information geographic information provided to users following the concepts of ubiquitous public access
[ISO 19154:2014]

ubiquitous public access UPA
service that enables end-users to have easy and interoperable access to specific types of data, irrespective of their location or access device, and that match their interest criteria
EXAMPLE Linked Geodata Service
NOTE 1 In the example, the Linked GeoData Service is responsible for openly inter-connecting geographic information to external repositories or web resources using a transform to either Resource Description Framework (RDF) or Web Ontology Language (OWL) format.
[ISO 19154:2014]

UML application schema application schema written in UML in accordance with ISO 19109
[ISO 19136-1:2020]

UML template parameterized model element that describes or identifies the pattern for a group of model elements of a particular type
[IBM Rational System Developer, ISO 19129:2009]

uncertainty parameter, associated with the result of measurement, that characterizes the dispersion of values that could reasonably be attributed to the measurand
NOTE 1 When the quality of accuracy or precision of measured values, such as coordinates, is to be characterized quantitatively, the quality parameter is an estimate of the uncertainty of the measurement results. Because accuracy is a qualitative concept, one should not use it quantitatively, that is associate numbers with it; numbers should be associated with measures of uncertainty instead.
[ISO 19116:2019]

ungeoreferenced grid gridded data that does not include any information that can be used to determine a cell's geographic coordinate values
EXAMPLE A digital photo without georectification information included.
[ISO 19163-1:2016]

Uniform Resource Identifier URI
unique identifier for a resource, structured in conformance with IETF RFC 2396
NOTE 1 The general syntax is <scheme>::<scheme-specific-part>. The hierarchical syntax with a namespace is <scheme>://<authority><path>?<query> - see RFC 2396.
[ISO 19136-1:2020]

unit defined quantity in which dimensioned parameters are expressed.
NOTE 1 In this International Standard, the subtypes of units are length units, angular units, time units, scale units and pixel spacing units.
[ISO 19111:2019]

unit of measure reference quantity chosen from a unit equivalence group
NOTE 1 In positioning services, the usual units of measurement are either angular units or linear units. Implementations of positioning services must clearly distinguish between SI units and non-SI units. When non-SI units are employed, it is required that their relation to SI units be specified.
[ISO 19116:2019]

universal representation universal feature model to be specified without knowing users' structures or abstraction models
[ISO 19154:2014]

universe of discourse view of the real or hypothetical world that includes everything of interest
[ISO 19101-1:2014]

unqualified cardinality <OWL> cardinality restriction that applies to all literals or individuals that are connected by a data property or an object property
[OWL, ISO 19150-2:2015]

user active object that initiates service requests to the system
NOTE 1 Users are usually objects that act as proxies for people accessing the functionality of the system.
[ISO 19132:2007]

utility network network describing the legal space of the topology of a utility
EXAMPLE The legal space needed to access and to keep in repair a cable or pipeline utility network.
NOTE 1 A utility network may be attributed with information about its legal, recorded or informal space.
NOTE 2 A utility network can also be modelled as a basic administrative unit.
[ISO 19152:2012]

valid time time when a fact is true in the abstracted reality
[ISO 19108:2002]

validation process of assessing, by independent means, the quality of the data products derived from the system outputs
[CEOS WGCV, ISO 19101-2:2018]

value <UML> element of a type domain
NOTE 1 A value may consider a possible state of an object within a class or type (domain).
NOTE 2 A data value is an instance of a data type, a value without identity
NOTE 3 A value can use one of a variety of scales including nominal, ordinal, ratio and interval, spatial and temporal. Primitive datatypes can be combined to form aggregate datatypes with aggregate values, including vectors, tensors and images.
[ISO/IEC 19501:2005, ISO 19156:2011]

value domain set of accepted values
EXAMPLE The range 3-28, all integers, any ASCII character, enumeration of all accepted values (green, blue, white).
[ISO 19103:2015]

vector quantity having direction as well as magnitude
NOTE 1 A directed line segment represents a vector if the length and direction of the line segment are equal to the magnitude and direction of the vector. The term vector data refers to data that represents the spatial configuration of features as a set of directed line segments.
[ISO 19123:2005]

vector geometry representation of geometry through the use of constructive geometric primitives
[ISO 19144-1:2009]

vehicle traveller
object subject to being navigated or tracked
NOTE 1 Includes pedestrians. See ISO 14825. In this International Standard, vehicle can be replaced by trav-

eller without any change of intent.
[ISO 19133:2005]

vehicle classification type of vehicle, based on the nature of its construction or intended purpose.
NOTE 1 Classifications based on construction include automobile, truck, bus, bicycle, etc. Classifications based on purpose include taxi, emergency vehicle, etc. Vehicle classification can be used to determine the application of navigation constraints.
[ISO 19133:2005]

velocity vector <Radar> first derivative of the antenna's position vector
[ISO 19130-2:2014]

verification provision of objective evidence that a given item fulfils specified requirements
NOTE 1 When applicable, measurement uncertainty should be taken into consideration.
NOTE 2 The item may be, e.g. a process, measurement procedure, material, compound, or measuring system.
NOTE 3 The specified requirements may be, e.g. that a manufacturer's specifications are met.
NOTE 4 Verification should not be confused with calibration. Not every verification is a validation.
[ISO/IEC Guide 99:2007, ISO 19159-1:2014]

verification test test developed to prove rigorously whether an IUT is correct
[ISO 19105:2000]

version (temporal) complete representation of an object at a given instance in time
NOTE 1 Temporal versions differ from samples in that a complete description is required. In this sense a version is a complete sample able to be considered outside the domain of the temporal sequence to which it may belong.
[ISO 19132:2007]

vertical accuracy measure of the positional accuracy of a dataset with respect to a specified vertical datum
[ISO 19159-2:2016]

vertical coordinate reference system one-dimensional coordinate reference system based on a vertical reference frame
[ISO 19111:2019]

vertical coordinate system one-dimensional coordinate system used for gravity-related height or depth measurements
[ISO 19111:2019]

vertical reference frame vertical datum
reference frame describing the relation of gravity-related heights or depths to the Earth
NOTE 1 In most cases, the vertical reference frame will be related to mean sea level. Vertical datums include sounding datums (used for hydrographic purposes), in which case the heights may be negative heights or depths.
NOTE 2 Ellipsoidal heights are related to a three-dimensional ellipsoidal coordinate system referenced to a geodetic reference frame.
[ISO 19111:2019]

vertical reference system VRS
set of conventions defining the origin, scale, orientation and time evolution that describes the relationship of gravity-related heights or depths to the Earth
NOTE 1 The abstract concept of a VRS is realised through a vertical reference frame.
[ISO 19111:2019]

vicarious calibration post-launch calibration of sensors that make use of natural or artificial sites on the surface of the Earth
[ISO 19159-1:2014]

viewpoint (on a system) form of abstraction achieved using a selected set of architectural concepts and structuring rules, in order to focus on particular concerns within a system
[ISO/IEC 10746-2, ISO 19154:2014]

vocabulary terminological dictionary which contains designations and definitions from one or more specific subject fields
NOTE 1 The vocabulary may be monolingual, bilingual or multilingual.
[ISO 1087-1:2000, ISO 19146:2018]

waypoint location on the network that plays a role in choosing candidate routes potentially satisfying a routing request
[ISO 19133:2005]

Web service service that is made available through the Web
NOTE 1 A Web service usually includes some combination of programming and data. It can also include human resources.
[ISO 19101-1:2014]

whiskbroom sensor sensor that sweeps a detector forming cross-track image line(s) and constructs a larger image from a set of adjacent lines using the along-track motion of the sensor's collection platform
[ISO 19130-1:2018]

white space sequence of one or more characters that have no glyphs
[ISO/IEC 9075-2:2016, ISO 19162:2019]

workflow automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules
[ISO 19119:2016]

World Wide Web Web
universe of network-accessible information and services
[ISO 19101-1:2014]

Index

H

Q