Requirements for data management and maintenance to support regional land use research

Henning Sten Hansen[1], Pia Viuf[1], Wolfgang Loibl[2], Jan Peters-Anders[2], Sergey Zudin[3], Jürgen Vogt[4]

(1) National Environmental Research Institute (NERI), University of Aarhus, Roskilde, Denmark
(2) Austrian Research Centers GmbH (ARCsys), Austria
(3) European Forest Institute, Joensuu, Finland
(4) Joint Research Centre (JRC), European Commission, Ispra, Italy

Abstract

SENSOR is dependent on sufficient reliable and accurate data that have to be provided and shared by the partners within the project. Access to reliable and harmonised data across Europe is a fundamental precondition for realisation of the SENSOR project. The current chapter describes basics concerning geo-spatial data types and formats, system architecture and database technologies, interoperability standards, including the INSPIRE principles, data warehouse and GeoPortal technologies. Further some information on spatial data mining, on data policies and related legal aspects and the SENSOR approach for spatial data handling are provided.

Keywords

Spatial data management, land use, INSPIRE, spatial data mining

1 Introduction

The quality of examining landscape related phenomena like sustainability impact assessment for landscape multifunctionality as achieved within SENSOR is dependent on sufficient reliable and accurate data that have to be provided and shared by the partners within the project. A proper geospatial data management and data sharing system including metadata reporting, data retrieval, data viewing, data upload and download is the backbone of landscape related research.

Geographic Information Systems are built using formal models that describe how objects are located in space. Every geographical object or phenomena can basically be represented by a point, line or polygon – plus some attributes describing the object. Geographical data are referenced to locations on, below or above the earth's surface by using a standard reference system. There are at least two fundamental different ways of representing geographic information: vector representation and raster representation (Figure 1).

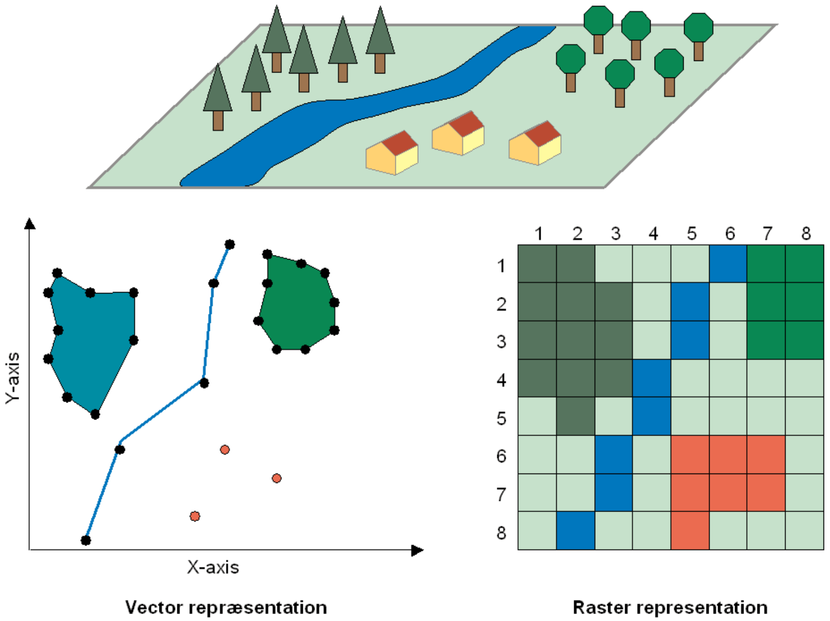

Fig. 1. Vector versus raster representation

Vector is a data structure, used to store spatial data. Vector data is comprised of lines or arcs, defined by beginning and end points, which meet at nodes. The locations of these nodes and the topological structure are usually stored explicitly. Features are defined by their boundaries only and curved lines are represented as a series of connecting arcs. Vector storage involves the storage of explicit topology, which raises overheads, however it only stores those points, which define a feature, and all space outside these features is 'non-existent'.

Raster is an alternative method for representing spatial data. Each area is divided into rows and columns, which form a regular grid structure. Each cell must be rectangular in shape, but not necessarily square. Each cell within this matrix contains location coordinates as well as an attribute value. The spatial location of each cell is implicitly contained within the ordering of the matrix, unlike a vector structure, which stores topology explicitly. Areas containing the same attribute value are recognised as such, however, raster structures cannot identify the boundaries of such areas as polygons. Within the SENSOR community we have to use as well vector as raster based spatial information. Generally, it can be troublesome to use a mixture of data models, but we have to rely on available data. You can transform the data from raster to vector and vice versa but generally not without loss in quality.

Geospatial data have both spatial and thematic properties. Conceptually, geographic data can be divided into two elements: entities and attributes. GIS have to be able to manage both elements, and this defines the overall requirements to the database technology behind.

We propose a definition of a spatial database system as a database system that offers spatial data types in its data model and query language and supports spatial data types in its implementation, providing at least spatial indexing and spatial join methods. Spatial database systems offer the underlying database technology for geographic information systems and other applications.

The Open Geospatial Consortium (http://www.opengeospatial.org/) is a global key player working for interoperability between various database systems. The requirements for implementations of spatial databases are described in the implementation specifications for SQL (Open Geospatial Consortium, 2005). This part of OpenGIS® Simple Features Access (SFA), also called ISO 19125, is to define a standard Structured Query Language (SQL) schema that supports storage, retrieval, query and update of feature collections via the SQL Call-Level Interface. Open Geospatial Consortium allows three different approaches: a) the normalised geometry schema, b) the binary geometry schema, and c) the geometric data type implementation. Thus the database software suppliers have three different ways of handling spatial data in (object-) relational database systems. Oracle, Informix and DB2 have all developed versions based on SQL with geometric data types. The open source databases PostgreSQL and MySQL have also developed versions with geometry data types. However, Microsoft SQL Server does not have spatial data types.

To allow merging and combining different Geospatial data a common coordinate system is required. The content of the curved surface of the Earth is transferred to a flat plane by a projection. Mapping of ellipsoidal and spherical coordinates to plane coordinates cannot be performed without distortion in a plane coordinate system. Distortion can be controlled, but not avoided. Various projections exist to perform such a transfer to reduce distortion in certain ways: among them conic projections (e.g. the Lambert projections), transverse cylindrical projections (e.g. Mercator projections) or plane coordinate projection.

The ellipsoid's properties describing size, shape, position and orientation is summarised as "Datum". To map entire Europe, today the European Terrestrial Reference System 1989 (ETRS89) is committed as the geodetic datum.

The ETRS89 Transverse Mercator Coordinate Reference System is recommended for pan-European mapping at scales larger than 1:500 000. For pan-European conformal mapping at scales smaller or equal 1:500 000 the ETRS89 Lambert Conformal Conic Coordinate Reference System is recommended. With conformal projection methods attributes such as area will not be distortion-free. For pan-European statistical mapping at all scales or for other purposes where true area representation is required, the ETRS89 Lambert Azimuthal Equal Area Coordinate Reference System (ETRS-LAEA) is recommended. The Lambert Equal Area projection is recommended for use in the SENSOR project.

2 Data Infrastructure: Principles of Distributed GIS Technology

Developing a common data infrastructure requires some degree of standardisation among the various data sets. Although, the standards of interest to the SENSOR project are not static but will evolve during the project period as technology changes, the draft specifications of the INSPIRE initiative on architecture, standards and metadata are the main guidelines for this task (INSPIRE, 2002 a). Based on this foundation, an overall frame for the data infrastructure including Web-based catalogue services enabling participants to discover and download appropriate data for their work will be designed and a prototype developed (Figure 2).

The main aim of the SENSOR Data Management System is to support the project partners concerning data handling. To do this the system will include the following components

- Data Warehouse
- Geoportal (Clearinghouse mechanism)
- Metadata reporting system
- Upload and download of data

- Pre- and post processing tools

Besides these IT components the SENSOR Data Management system contains a defined set of Core data and a SENSOR Data Policy.

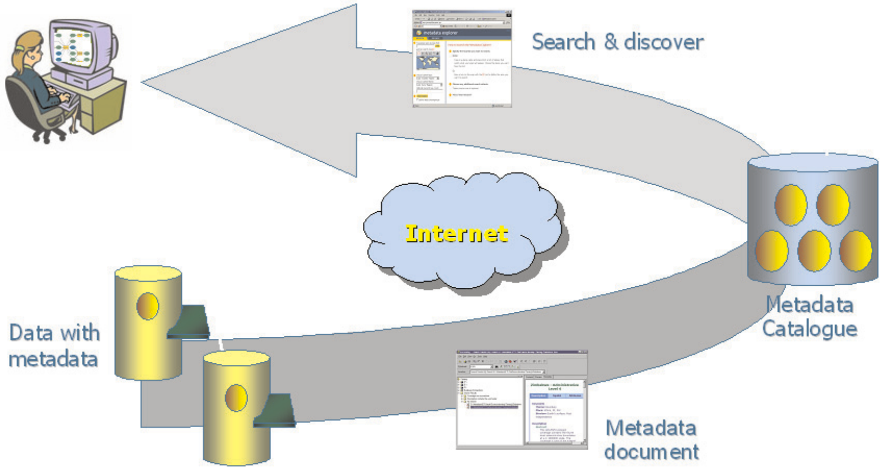

Fig. 2. The Data Management System from a user's point of view.

GIS technology is evolving beyond the traditional GIS community and becoming an integral part of the information infrastructure in many organisations. The unique integration capabilities of a GIS allow disparate data sets to be brought together to create a complete picture of a situation. Thus organisations are able to share, coordinate, and communicate key concepts among departments within an organisation or among separate organisations using GIS as the central Spatial Data Infrastructure. GIS technology is also being used to share information across organisational boundaries via the Internet and with the emergence of Web services. However, other obstacles like for example lack of semantic interoperability may impede the use of information.

An open GIS system allows for the sharing of geographic data, integration among different GIS technologies, and integration with other non-GIS applications. It is capable of operating on different platforms and databases and can scale to support a wide range of implementation scenarios from the individual consultant or mobile worker using GIS on a workstation or laptop to enterprise implementations that support hundreds of users working across multiple regions and departments. An open GIS also exposes objects that allow for the customisation and extension of functional capabilities using industry standard development tools. The current chapter will describe some of the most important elements of distributed GIS, as we will use the concept in SENSOR.

2.1 Standards and Interoperability

Interoperability and open architectures are core requirements for state of the art implementations of IT solutions (Klopfer, 2006). Service oriented architectures based on a commitment to use open standards enables a system of component based building blocks, which can be chosen, run and maintained according to their best match of user requirements, independent of vendor solutions or storage models.

Standards define the common agreements that are needed to achieve interoperability between IT components (Figure 3). Standardisation bodies like ISO or CEN are developing de jure standards, whereas organisations like the Open Geospatial Consortium (OGC) develops specifications that by a consensus process and their common acceptance become de facto standards. Several ISO TC/211 standards are of high importance for building Spatial Data Infrastructures. Besides the ISO Standards the Open Geospatial Consortium (OGC) has developed implementation rules to ensure interoperability. Products and services compliant to OpenGIS interface specifications enable users to freely exchange and apply spatial information, applications and services across networks, different platforms and products.

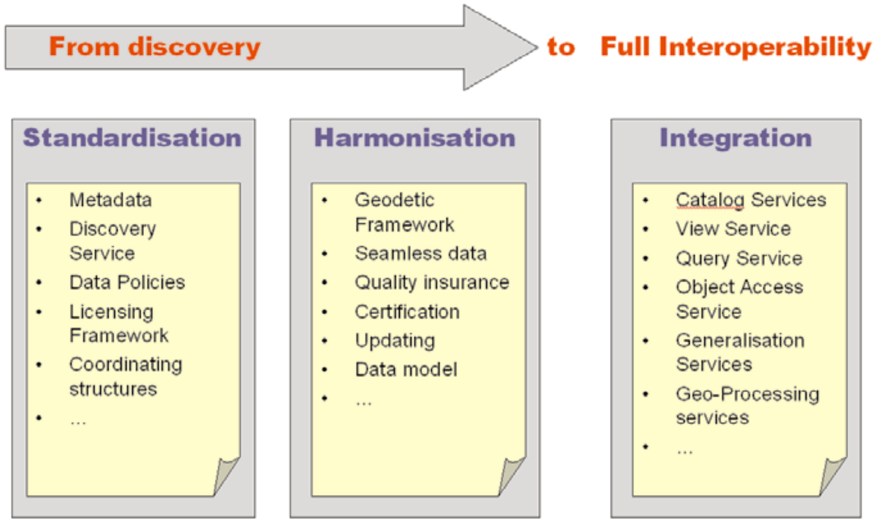

Fig. 3. Ways towards a spatial information infrastructure (INSPIRE, 2002c)

Besides these GI related standards, a geospatial data infrastructure is built on general IT standards like XML (extensible Mark-up Language, SOAP (Simple Object Access Protocol) and WSDL (Web Services Description Language). This is important because GI systems are not longer isolated stand alone systems, but nowadays integrated in the general IT infrastructures. Besides, a basic foundation for all data related work in an EU-funded project like SENSOR is the draft INSPIRES principles (INSPIRE, 2002 a).

2.2 Data Warehouse Architecture

A Data Warehouse is defined as a subject-orientated, integrated, time-variant, non-volatile collection of data that support the decision-making process in an organisation (ESRI, 1998). In general a Data Warehouse is a large database organising data from various sources in a repository facilitating query and analysis. The database has to be structured and contain key data, for search and retrieval. The spatial data warehouse in SENSOR responds to several needs. First we have to realise that the SENSOR project involves 35 partners from many countries, and the data sources are very widely spread. The main task for the central database is to facilitate access to data for all partners. Most common data sets should be added to the Data Warehouse and harmonised so they match with the overall system architecture and the geo-reference characteristics and data quality standards. Data downloaded from EuroStat, ESPON, or the European Environment Agency are not usable at once, but must be adapted in various ways – first of all due to differences in the database keys used.

2.3 GeoPortals and Clearinghouses

Efficient use of geographic information assumes access to documentation that describes origin, quality, age, ownership and suitability for certain purposes. This associated information is referred to as metadata (see paragraph 2.4). A key component of any spatial data infrastructure is a catalogue with metadata that can be used in searching for data considering geometric data content, geographic location, time and thematic attributes.

Technically the word portal refers to a web site acting as an entry point to other web sites (Tait, 2005). An extended definition of a GeoPortal will be a web site that represents an entry point to sites with geographic content. Spatial portals were developed as gateways to SDI initiatives and served as contact point between users and data providers. The GeoPortal allow users to search and browse between huge amounts of data. One of the earliest attempts to develop a Geoportal was the US Federal Geographic Data Committee's Clearinghouse, in Europe the INSPIRE proposal resulted in the development of a European Geoportal (Bernard et al., 2005).

Geoportals can be divided into two groups: Catalogue Geoportals and Application Geoportals (Tang and Selwood, 2005). Catalogue portals create and maintain indexes describing available information services. Catalogue portals are useful when they provide information to a wide variety of services, data providers and user groups. Application portals combine information services into a Web based mapping application that generally focuses on a particular task. Their target community is well defined and they provide efficient access to data and functional services, which the portal manager selects to meet the user's needs. In the SENSOR project a combination between Catalogue and Application Portal is used in order to support both the data and the application side.

The publishing process is the most important part – without any metadata it is impossible to carry out a proper search for data. Publishing comprises addition, modification and deletion of metadata. The SENSOR project has focused much on this effort and a web based metadata publishing / reporting system has been available since August 2005.

Geoportals are built using the World Wide Web infrastructure technology and GIS software. The front end typically sits on top of an Internet Map Server that delivers the services. A Geoportal contains three components: Web Portal, Web services and Data Management. Table 1 describes the components, their relationships between each other and the standards and technologies they are built upon.

Table 1. Geoportal architecture (After Tait, 2005)

Components	Elements	Environments	Functions
Web Portal	Web site	HTML, HTTP, XML, XSL, JSP, ASP	Search, View, Publish, Admin.
	Web controls	Java beans, .NET	Query, Map, Edit
Web services	Geo Web services	XML, SOAP, WSDL, WMS, WFS, GML	QUERY, Render, Transaction
Data Management	RDBMS		Vector
		SQL	Raster
	Data		Tabular

2.4 Metadata

Electronic searching and exchange of metadata require standardisation. Metadata must follow the ISO 19115 standard for metadata. Since 1994, ISO/TC211 (http://www.isotc211.org/) has been working to establish a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth. ISO 19115 Geographic Information - Metadata is a part of the family of TC211 standards and it defines the term METADATA as "data about data". The objective of ISO 19115 is to specify a structure for describing digital geographic data, and the ISO standard on geographic information has quite recently been adopted by CEN the European Standard Organisation.

SENSOR consortium realised from the very beginning, that in order to build a strong spatial data infrastructure and to establish integrity and consistency of all data, metadata would be crucial. Metadata and metadata servers enable users to integrate data from multiple sources, organisations, and formats. Metadata for geographical data may include the data source, its creation date, format, projection, scale, resolution, and accuracy.

Due to the fact that SENSOR end user is the European Commission, it seems reasonable to take outset in existing metadata standards within the Commission. At first, we therefore took a look on the metadata profile from the EEA (European Environmental Agency) as an initial metadata set. The EEA metadata profile builds on the principles in ISO 19115 as well as INSPIRE. Currently, a Metadata Core Drafting team is working on a detailed metadata specification for INSPIRE. The attribute set was reduced for SENSOR in order to increase acceptance among SENSOR data deliverers to fill out the forms completely. The metadata set shall fulfil all needs within the project to fully inform all team members about the content of the data sets. The metadata is furthermore a precondition to assess the usability of the respective data.

Therefore some additional attributes, which are not considered by ISO 19115- standard, but seem to be important, have been added. The most important among them are the fields containing thematic statistical information (e.g. demographic or economic data based on administrative entities like NUTS-Regions) and further content regarding spatial characteristics (e.g. land use classes, elevation, terrain shape, environmental pollution).

Metadata for single data sets can be stored within a metadata-XML-file via ESRI’s ArcCatalog in several different style sheets: among them the US-standard “FGDC” structure and the “ISO”-style sheet. Despite the guideline to use ISO-structure we recommend to use the ArcCatalog default metadata editor and FGDC style sheet, which considers FGDC structure, as this is the only way to allow storage of attribute data information (ISO has the disadvantage that no attribute information integration is available within this structure). We apply a special ISO-metadata version, by allowing information about the attribute fields (name and data type). The metadata copied to the SENSOR metadata base can be overwritten at any time. Thus it is recommended to obtain the metadata first from XML-files – for easily including of geodetic information - and to correct or extend the entries afterwards if necessary.

3 SENSOR Data Management System Design

The overall objective of the SENSOR Data Management System is to support all partners to get access to data from various sources as well as data produced within the SENSOR project (Figure 4). The first element in the SENSOR Data Management System is the Metadata Publishing System aimed at reporting metadata for all data related to the SENSOR project. Parallel to the metadata reporting the application facilitates the upload of data to the central server. Closely related to the upload procedure is a checking tool for tabular data regarding geo-reference code (frequently a NUTS-code). Finding and discovering data is provided through a retrieval system based on metadata keywords for the entire data set collection and provided through ESRI's Metadata Explorer.

3.1 SENSOR Data Warehouse

The main component in the SENSOR Data Management system is the Data Warehouse storing pre-processed spatial data with associated metadata (fig. 4). All common data used in the SENSOR project as well as all data produced by SENSOR will be available from the Data Warehouse.

The Data Warehouse is based on state-of-the-art database technology using ArcSDE 9.2 from ESRI - providing a gateway for storing, managing, and accessing spatial data in any of several leading RDBMS from any ArcGIS application. It is a key component in managing a shared, multi-user Geodatabase in a RDBMS. Currently ArcSDE supports the following relational databases: Oracle, IBM DB2 Universal Database, IBM Informix Dynamic Server, and Microsoft SQL Server. Within SENSOR the underlying relational database system will be Microsoft SQL Server 2005.

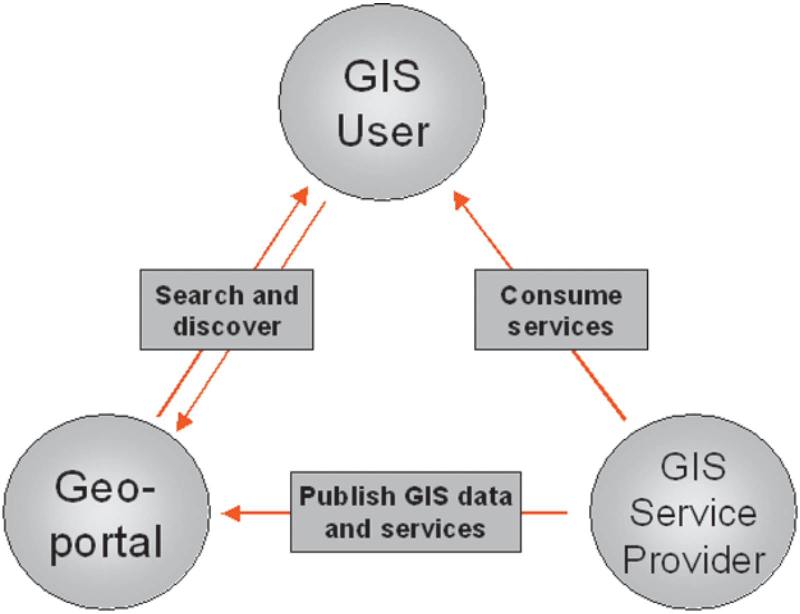

Fig. 4. The relationships between Geoportal, user and service provider (After Tang & Selwood 2005).

Spatial data in the SENSOR project are stored in ArcSDE as either vector features or as raster data sets along with traditional tabular attributes. Topology – the spatial relationships between geographic features – is fundamental to ensuring data quality (ESRI, 2005; Silvertand, 2004). Topology in ArcSDE is implemented as a set of integrity rules that define the behaviour of spatially related geographic features and feature classes. Topology is used to manage the integrity of coincident geometry between different feature classes – for example to check if the coastlines and country boundaries are coincident. The various components of the SENSOR Data Management System (Figure 5) is further described below.

3.2 Input – The SENSOR Metadata Publishing and Upload Application

The SENSOR Metadata Publishing System was developed as Web based Java application as the first part of the SENSOR Data Management System. The purpose was to give the SENSOR community tools for uploading various NUTS-related data and generic geospatial data to the Data Management System. To ensure high convenience for metadata upload the system has a graphical user interface (GUI) guiding the user easily through the application offering several forms with pull down menus just to click an entry among alternatives and some free entry fields for individual text.

A final upload of geospatial data or tabular data with spatial reference is possible only after all metadata have been entered completely. Additional features of this application are automated data integrity checking and a (preliminary) basic data retrieval tool tracing the metadata of the entire geospatial data collection for certain files related to certain keywords.

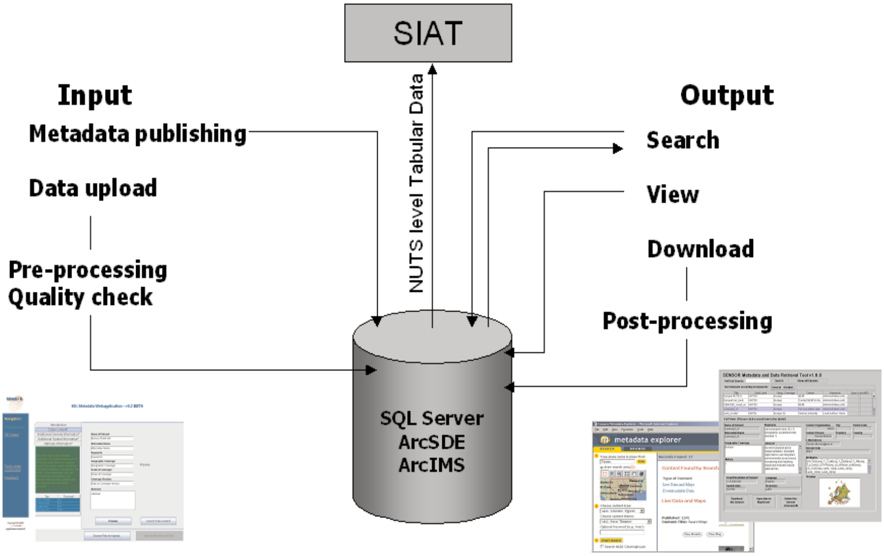

Fig. 5. Principles for SENSOR Data Management System

3.3 Output – The SENSOR Geoportal

The general entrance to the system is through the SENSOR GeoPortal. The Geography Network Explorer as well as the INSPIRE GeoPortal are both examples on how to use an Internet Map Server based Geoportal for searching, discovering and retrieval of data. The SENSOR GeoPortal is based on an Internet Map Server, and the main competitors among Internet Map Servers are ArcIMS from ESRI and MapServer, which is an Open Source implementation. MapServer is free of charge and an open concept, with unlimited possibilities for developing targeted implementations. This is obviously an advantage. However, the implementation effort can be a rather tough job, because we have to develop the whole end user application by ourselves, and this is certainly a disadvantage. ArcIMS is a rather expensive product, but comes with built-in applications for administration and authoring as well as end user applications. However, we still have the possibility to extend the standard applications – or even build our own application using Java. The choice between the two alternatives is at first sight not easy, but taking into account that the software environments of the end users are ESRI based it was decided to use ArcIMS.

The OGC WMS connector produces maps of geo-referenced data in image formats (PNG, GIF, JPEG) and creates a standard means for users to request maps on the Web and for servers to describe data holdings. The OGC WFS connector enables ArcIMS to provide Web feature services that adhere to the OpenGIS Web Feature Service Implementation Specification. The connector provides users with access to geographic (vector) data, supports query results, and implements interfaces for data manipulation operations on Geographic Mark-up Language (GML) features served from data stores that are accessible via the Internet. GML is an OpenGIS Implementation Specification designed to transport and store geographic information, and it is a encoding of Extensible Mark-up Language. The main development environments for the SENSOR Geoportal are Java and ArcXML, which is the protocol for communicating with the ArcIMS Spatial Server (ESRI, 2002).

3.4 Spatial Data Mining

The immense amount of geographically referenced data produced by developments in digital mapping, remote sensing, and the global diffusion of GIS emphasises the importance of developing *data driven* inductive approaches to geographical analysis and modelling to facilitate the creation of new knowledge and aid the processes of scientific discovery (Openshaw, 1999). Spatial data mining aims to uncover spatial patterns and relations.

The main difference between data mining in relational database systems and in spatial database systems is that attributes of the neighbours of some object of interest may have an influence on the object and therefore have to be considered as well (Ester et al. 2001). The explicit location and extension of spatial objects define implicit relations of spatial neighbourhood (such as topological, distance and direction relations), which are used by spatial data mining algorithms. Therefore, new techniques are required for effective and efficient data mining.

There are several major categories of data mining techniques (Ester et al., 1997):

- **Clustering** is the task of grouping objects into meaningful subclasses, so that members of a cluster are as similar as possible, whereas the members of different clusters differ as much as possible from each other. Thus clustering can be used to discover regions with low economic growth.
- **Characterisation** is the task to find a compact description for a selected subset of objects – e.g. to characterise certain target regions such as areas with a high percentage of unemployed. Spatial characterisation does not only consider the attributes of the target regions but also neighbouring regions and their properties.
- **Classification** refers to the task of discovering a set of classification rules that determine the class of any object form the values of its attributes.
- **Spatial trends** describe a regular change of non-spatial attributes when moving away from certain start objects. Global and local trends can be distinguished. To detect and explain such spatial trends, e.g. with respect to the economic power, is an important issue in geography.

A major challenge for this part of the SENSOR Data Management implementation is therefore to do research and development on effective methods for determining spatial and non-spatial relationships between datasets. The tools are based on recent advances in spatial data mining and knowledge discovery as described by Ester, Kriegel and Sander (2001) and facilitate *location prediction*, *spatial association*, *spatial clustering* and *spatial trend detection*.

4 Data policy

The data policy covers aspects of data access, ownership, licensing, and Intellectual Property Rights (IPR) on the data used within the framework of the SENSOR project. The SENSOR data policy follows the principles to be developed under the INSPIRE initiative (INSPIRE, 2002b). Currently, however, only a position paper on ‘Data Policy and Legal Issues exists, which lacks relevant details. As a consequence the SENSOR data policy has been developed as a consensus among the SENSOR partners, following the indications given in the INSPIRE position document. It might need revision when more detailed guidelines become available under the INSPIRE initiative.

Following these principles, it will be important that all data used and generated in the frame of SENSOR are well documented following strictly the SENSOR metadata profile (Table 2) and that the relevant search facilities are available. Furthermore, it is important that all data are available to the whole SENSOR community under clear conditions. Questions of data ownership, copyrights and conditions have now been clarified in order to encourage the disclosure and upload of data available as well as their widespread use within the SENSOR community.

4.1 Upload policy

All partners are encouraged to upload metadata on data of common interest and possibly to upload the data themselves. The uploading institution will retain the ownership of the data and will specify the conditions of use of the data. For any dataset to be uploaded, a copyright statement must be included in the metadata. By uploading the data, the data provider (owner) agrees that all SENSOR partners have free access to the data for their work within the SENSOR project. If not explicitly specified otherwise, all other uses will have to be authorised. It is strictly forbidden to deliver data to third parties outside the SENSOR project or to use the data for purposes outside the SENSOR project without the written consent of the data owner. Inquiries from third parties should be transferred to the data owner for clarification. All datasets must be accompanied by metadata, and the metadata will be freely available also for further (public) distribution. Data sets can be uploaded once the metadata are completely available and the data policy and copyright agreement has been accepted.

Table 2. The metadata list for SENSOR with associated ISO 19115-Standard codes

Metadata on metadata		ISO Code
• Point of contact		
• Name of contact organisation	*	8.376
• Name of contact person	*	8.375
• Address: City		8.378.389.382
• Address: Province, state		8.378.389.383
• Address: Postal code		8.378.389.384
• Address: Country		8.378.389.385
• Address: E-mail	*	8.378.389.386
• Address: web link	*	SENSOR specific

Data set identification

• Title of the data set	*	15.24.361
• Abstract	*	15.25
• Keywords	*	15.33.53
• Topic category	*	15.41
• Date of version	*	15.24.362.394

Reference system (SENSOR: information transferred via XML)

• Name of reference system	(*)	13.196.207
• Datum name	(*)	13.192.207
• Projection (Information via XML)		
• Name of projection	(*)	13.190.207
• Standard parallel	(*)	13.194.217
• Longitude of central meridian	(*)	13.194.218
• Latitude of projection origin	(*)	13.194.219
• False easting	(*)	13.194.220
• False northing	(*)	13.194.221
• False easting northing units	(*)	13.194.222
• Scale factor at equator	(*)	13.194.223
• Longitude of projection centre	(*)	13.194.224
• Latitude of projection centre	(*)	13.194.225

Distribution information

• Owner		
• Name of owner organisation	*	15.29.376

Other information

• Language within the data set	*	15.39
• Exchange format		
• Name of exchange format	*	15.32.285
• Scale	*	15.38.60.57
• Resolution (if raster data set)	SENSOR specific	
• Spatial Entities (NUTS-hierarchy)	SENSOR specific	
• Data type (vector / raster / tabular)	SENSOR specific	
• List of attributes	SENSOR specific	

(Attribute information via XML-file or XLS-table-header –parsing)

The "Data-type"- line above indicates, that our metadata profile is not only considering geo-spatial data but is also dealing with tabular data, which are non-spatial but referenced to spatial entities via identification code (ID).

4.2 Download policy

All SENSOR partners have full access to the metadata system, where they can search for data and information on the conditions of their use. Available datasets can be downloaded for use within the SENSOR project. Before downloading the data, the user agrees on the conditions of use of the data (data policy and copyright agreement).

4.3 SENSOR accepted Data formats

Data submitted to the Data Management System should follow certain standards. XML is emerging as the international standard for exchange of information, and you can easily import and export XML data in most modern GI software systems like ArcGIS. However due to the often huge size of geographic data sets, XML has had limited success in the GI Community. Instead native data formats from vendors like ESRI are used. In the SENSOR project, data should be exchanged in one of the following formats:

1) ESRI Shapefiles;
2) ESRI Personal Geodatabases;
3) Erdas Imagine or TIFF;
4) ESRI Coverages and Grids via Exchange File Format (E00);
5) XML / GML;
6) Tabular data (e.g., statistics for administrative regions).

These data need to be linked to a geographic entity via a common feature code – often a NUTS identification.

In principle, SENSOR data should comply with INSPIRE recommendations. This implies that data should be provided in a compliant reference and projection system, i.e. ETRS89 specifications (Annoni et al., 2003) and that grids should follow the INSPIRE grid specifications (JRC, 2003). This is very important in order to make these data readily available and useable for different applications. In case partners should have problems to convert the data, the data management team can try to help to solve the problem, provided that the data provider is able to give a detailed and accurate description of the projection system of the data. However, we underline that this should not be the rule and that in principle it remains the task of the different modules to provide data in the correct projection system.

5 Core data

The INSPIRE Working Group on Reference Data and Metadata encourages establishing a reference or core data set as an instrument to harmonise data from various sources. The recommendations from this group were: a) Geodetic reference data; b) Units of administration; c) Units of property rights (parcels, buildings); d) Addresses; e) Selected topographic themes (hydrography, transport, height); f) Orthoimagery; g) Geographical names.

During the further work with INSPIRE, the reference data set was changed a little bit – now also including European Grid in the so-called annex 1 data (COM, 2004). Within SENSOR we have chosen a geodetic reference system, administrative boundaries in form of NUTS, European Grid, CORINE Land cover, LANMAP and the European Digital Elevation model as our reference data set. By defining a SENSOR core data set we encourage partners to use for example the same NUTS map – although many different versions are available. Concerning the role as data harmonisation element, the datum, the projection, the NUTS administrative boundaries and the European Grid play the most important role. Those are described below.

5.1 NUTS

EuroStat established the Nomenclature of Territorial Units for Statistics (NUTS) more than 25 years ago in order to provide a single uniform breakdown of territorial units for the production of regional statistics for the European Union. The NUTS classification has been used since 1988 in Community legislation. But only in 2003, after 3 years of preparation, a Regulation of the European Parliament and of the Council of NUTS was adopted. From 1st May 2004, the regions in the 10 new Member States have been added to the NUTS.

The NUTS nomenclature is currently defined only for the 27 member states of the European Union. For additional countries comprising the European Economic Area (EEA) and also for Switzerland, a coding of the regions has been accomplished in a way, which resembles the NUTS. The NUTS map in SENSOR is based on SABE (Seamless Administrative Boundaries in Europe), which is an official product developed by EuroGeographics. The data behind SABE is the official administrative boundaries prepared by the national mapping agencies. The scale is generally 1:100,000. Aiming at a more equal size of the NUTS-3 polygons, within SENSOR a special version called NUTSx has been developed, where NUTS-3 is the basic map, but some countries, which have very small NUTS-3 entities - e.g. Germany - is represented by NUTS-2 (Renetzeder et al., 2008).

5.2 European Grid

The European grid should be used mainly for European purposes, but it can be useful also for national purposes. The datum to be used is ETRS89 as previously identified by INSPIRE. The geographical location of the grid points is based on the Lambert Azimuthal Equal Area coordinate reference system (ETRS-LAEA). The cartographic projection is centred on the point N 52°, E 10°. The coordinate system is metric. The square shape will appear when used in the defined projection, smaller or larger distortions will appear when re-projected to other projections.

Naming the individual cells can be done in several ways, but in SENSOR we have decided to use the so-called Direct Coordinate Coding System (DCCS), which concatenates the coordinates of Easting and Northing of a grid point. The length of the coordinates defines the precision of the grid. A grid with a precision of 1 m would require a maximum of 7 digits by each dimension. The resulting code would have 14 digits. A grid with a precision of 1 km would be defined by a code comprising 8 digits. Leading zeros are coded in order to preserve the precision information. Grid code identifies south-western corner of a cell.

5.3 CORINE Land cover

CORINE Land Cover (CLC) is a map of the European Environment Agency produced for the years 1990 and 2000. It provides comparable digital maps of land cover for each country for much of Europe (European Environment Agency, 1999, Bossard et al., 2000). The European land surface is classified using 44 classes of the 3-level CORINE nomenclature. CORINE Land Cover is produced mainly from satellite images, but aerial photos and near-ground imaging were also involved in the production. CORINE Land Cover in vector or raster formats is publicly available at no cost through the European Environment Agency web site.

5.4 LANMAP2

LANMAP2 is a Landscape Map at a scale of 1:2,000,000 covering the whole of Europe, from Iceland in the northwest to Azerbaijan in the southeast and from Gibraltar in the southwest to Novaya Zemlya in the northeast (Mücher et al., 2003; Mücher et al., 2006). Thus, LANMAP2 covers an area of approximately 11 million km^2. LANMAP2 is a hierarchical classification with four levels. The highest level (1) of the classification is determined by climate and has only eight classes. The second level is determined by climate and topography and has 31 classes. The third level, determined by climate, topography and parent material has 76 classes. In addition to this, the most detailed level (4) incorporates land cover and ends up with 350 landscape types.

6 Conclusion

The SENSOR Data Management System provides state-of-the-art core functionality for uploading data and metadata, storing data, searching and exploring data, selecting and downloading data. Use of off the shelf software complying with international standards like W3C, ISO TC/211 and the OGC are the implementation platform. When we talk about SENSOR Data Management we actually mean SENSOR Spatial Data Infrastructure dealing with all aspects of data management. Thus not only the technical aspects are included but also the economic and legal dimensions of data are addressed.

The first part of the Data Management System was already developed during summer 2005. This first component comprises the SENSOR Metadata Publishing system, and closely related to this is the data upload application, which still is under improvement. This data upload application could play an important role in establishing, at some level, data harmonisation and integrity.

The second part of the system was the implementation of the Data Warehouse with attached SENSOR GeoPortal for searching, exploring, selecting and downloading data. During this second phase, the connections between the Data Management system and SIAT have been established.

The third part of the system will deal with the development of tools for Spatial Data Mining and necessary pre- and post-processing tools. Data mining has the potential to equip users with extended analytical capabilities that can enable them to discover non-obvious relationships *between datasets*. By augmenting data discovery tools with spatial data mining, it is envisaged that users will discover related datasets that they would have otherwise overlooked. A major challenge for this final part of the SENSOR Data Management implementation is therefore to implement do research and development on effective methods for determining spatial and non-spatial relationships between datasets.

Generally speaking, during the process of developing the overall design of the SENSOR Data Management system, some "working" prototypes of different parts mentioned above have been developed. We see the main task for the nearest future in the bringing the various components together and establishing the integrated system.

References

Annoni A, Luzet C, Gubler E, Ihde J (eds)(2003): Map Projections for Europe. EUR 20120 EN (http://www.ec-gis.org/document.cfm?id=425&db=document)

Bernard L, Kanellopoulos I, Annoni A and Smits P (2005) The European geoportal—one step towards the establishment of a European Spatial Data Infrastructure. Computers, Environment and Urban Systems, 29, 15 – 31

COM (2004) Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL establishing an infrastructure for spatial information in the Community (INSPIRE)

European Environment Agency (1999) CORINE Land Cover – a key database for European integrated environmental assessment

Bossard M, Feranec J, Otahel J (2000) CORINE land cover technical guide – Addendum 2000. Technical report, 40. Copenhagen (European Environment Agency), p. 105. http://terrestrial.eionet.eea.int

ESRI (1998) Spatial Data Warehousing – An ESRI White Paper. Redlands, USA

ESRI (2002) ArcXML Programmers Reference Guide. Redlands, USA

ESRI (2005) ArcGIS 9 – Building a Geodatabase. Redlands, USA

Ester M, Kriegel HP, Sander J (1997) Spatial Data Mining: A Database Approach. Lecture Notes in Computer Science, vol. 1262, 47 – 66

Ester M, Kriegel HP and Sander J (2001) Algorithms and Applications for Spatial Data Mining. Geographic Data Mining and Knowledge Discovery, Research Monographs in GIS, Taylor & Francis

INSPIRE (2002 a) INSPIRE Architecture and Standards Working Group Position Paper. October 2002. http://inspire.jrc.it

INSPIRE (2002 b) INSPIRE Data Policy and Legal Issues Working Group Position Paper, October 2002. http://inspire.jrc.it

INSPIRE (2002c) INSPIRE: Reference Data and Metadata Position Paper, http://inspire.jrc.it

JRC (2003) Proceedings of the 1st European Reference Grid workshop and Proposal for a European Reference Grid coding system, 15-10-2004, JRC ESDI Action. (http://inspire.jrc.it)

Klopfer M (ed) (2006) Interoperability & Open Architectures: An Analysis of Existing Standardisation Processes & Procedures. OGC White Paper, Open Gepspatial Consortium 2006

Mücher CA, Bunce RGH, Jongman RHG, Klijn JA, Koomen AJM., Metzger MJ, and Wascher DM (2003) Identification and characterisation of environments and landscapes in Europe. (Alterra-rapport 832), Alterra, Wageningen (NL), 119pp

Mücher CA, Wascher DM, Klijn JA, Koomen AJM and Jongman RHG (2006) A new European Landscape Map as an integrative framework for landscape character assessment. In: Landscape Ecology in the Mediterranean: inside and outside approaches, Bunce RGH and Jongman RHG (eds) 2006. Proceedings of the European IALE Conference 29 March – 2 April 2005 Faro, Portugal IALE Publication Series 3, 233-243

Open Geospatial Consortium (2005) OpenGIS® Implementation Specification for Geographic information - Simple feature access - Part 2: SQL option

Openshaw S (1999) Geographical data mining: key design issues. Proc. Conference on GeoComputation, http://www.geovista.psu.edu/sites/geocomp99/Gc99/051/gc_051.htm

Renetzeder C, van Eupen M, Mücher S, Wrbka T (2008) Clustering Europe: a spatial regional reference framework for land use assessment. In: Helming K, Tabbush P, Perez-Soba (eds) Sustainability Impact Assessment and land use changes. Springer, 249-268

Silvertand G (2004) Storing and Maintaining Topology in an ArcSDE Geodatabase.. ArcUser July – September 2004, p 36

Tait MG (2005) Implementing Geoportals: Applications of Distributed GIS. Computers, Environment and Urban Systems, vol. 29, 33 – 47

Tang W and Selwood J (2005) Spatial Portals – Gateways to Geographic Information. ESRI Press