Chapter 5
Integrating and Generalising Volunteered Geographic Information

Monika Sester, Jamal Jokar Arsanjani, Ralf Klammer, Dirk Burghardt and Jan-Henrik Haunert

Abstract The availability of spatial data on the web has greatly increased through the availability of user-generated community data and geosensor networks. The integration of such multi-source data is providing promising opportunities, as integrated information is richer than can be found in only one data source, but also poses new challenges due to the heterogeneity of the data, the differences in quality and in respect of tag-based semantic modelling. The chapter describes approaches for the integration of official and informal sources, and discusses the impact of integrating user-generated data on automated generalisation and visualisation.

M. Sester (✉)
Institute of Cartography and Geoinformatics, Leibniz Universität Hannover, Hannover, Germany
e-mail: monika.sester@ikg.uni-hannover.de

J. Jokar Arsanjani
GIScience research group, Institute of Geography, University of Heidelberg, Heidelberg, Germany
e-mail: jokar.arsanjani@geog.uni-heidelberg.de

R. Klammer · D. Burghardt
Institute of Cartography, Dresden University of Technology, Dresden, Germany
e-mail: Ralf.Klammer@tu-dresden.de

D. Burghardt
e-mail: Dirk.Burghardt@tu-dresden.de

J.-H. Haunert
Institute of Geoinformatics and Remote Sensing, University of Osnabrück, Osnabrück, Germany
e-mail: janhhaunert@uni-osnabrueck.de

D. Burghardt et al. (eds.), *Abstracting Geographic Information in a Data Rich World*, Lecture Notes in Geoinformation and Cartography, DOI: 10.1007/978-3-319-00203-3_5,

5.1 Introduction

Data acquired by laymen, so-called Volunteered Geographic Information (VGI), and data from geosensor networks has lead to an increased availability of spatial information. Whereas until recently, authoritative data sets were dominating, VGI extends and enriches these data in terms of thematic variation and also by the fact that it is more user-centric. The latter is especially true for VGI collected by social media.

There are several reasons for integrating data from different sources rather than just relying on a single one: The heterogeneity of data models, sensors and application areas lead to complementary (and redundant) data covering our environment. Integration results in a higher completeness concerning geometric, temporal and thematic coverage, and quality. Especially the availability of sensors such as GPS-enabled devices and smart phones, leads to a plethora of new, previously unknown, data with various spatial characteristics: geo-located Flickr images, Tweets with location tags, Wikipedia articles, GPS-traces, just to name a few. In general, VGI acquisition can be distinguished between participatory and opportunistic. The latter refers to data, which users acquire more or less unconsciously, mostly with a non-specific purpose. A prominent example is the exploitation of mobile phone data to determine traffic information such as traffic jams. Participatory data acquisition requires the conscious participation of the user, as it is undertaken e.g. in the OpenStreetMap project, where users have the common goal to build up a map of the world in a joint, cooperative effort.

The general goal is the integrated and consistent use and visualisation of all sorts of available (spatial) data. The specific challenges posed by user-generated content are among others, the huge mass of data, which have to be handled and turned into valuable information. Other challenges are:

- The data are "as is"; often it consists of "unimportant" information chunks, and is of heterogeneous quality.
- The data are highly redundant, as many users may observe the same phenomenon. The challenge is to exploit and remove the redundancy in order to present a consistent result.
- The data are highly temporarily volatile: temporal analysis is needed in order to decide about the relevance of a chunk of information.
- The data are of variable, sometimes unknown explicit semantics, thus the thematic content has to be inferred from contextual information.
- The data has no explicit scale, although it can be said that most of the data is of large scale (e.g. Points of Interest—POIs); thus, the scale of the data has to be determined in order to integrate and visualise it.
- Often the data are raw data: in order to exploit it, it has to be analyzed and enriched with contextual information or by automatic interpretation (e.g. images of Flickr).

The goal is to provide integrated information with clear semantics, seamless geometry and consistent topology. There should be no visual conflicts, which would disturb and confuse the user—and possibly lead to incorrect decisions. Furthermore, the information should be carefully 'thinned' in order to provide no visual clutter. Thus, generalisation has to be applied. Finally, the heterogeneous nature and quality of the data should be made transparent to the user.

5.2 The Potential and Characteristics of User-Generated Content

In the following sections, two major sources of novel geographic information data are briefly sketched, namely geosensor networks and VGI.

5.2.1 Overview of Characteristics of Geosensor Network Data

A geosensor network is a network of geo-sensors, capable of measuring, communicating and calculating (Stefanidis and Nittel 2004). With these characteristics, it is possible that a geosensor network is not just a collection of sensors, but is able to develop some intelligence which allows it to solve complex tasks. Geosensor networks composed of a huge mass of possibly miniaturised sensors are able to sense the environment at unprecedented resolution. Matt Duckham (2012) called this capability "close sensing".

There are many application areas for geosensor networks, especially in environmental modelling, where other sensors are not available or cannot measure with the required sensitivity and resolution. Prominent examples are the observation of rare animal species, observation and monitoring of natural phenomena, risks (such as landslides, earthquakes), catastrophes (oil spills, fire evacuation), or traffic monitoring and control, to name but a few.

The underlying idea is that the sensors not just measure and communicate the data to a central server, where all information is aggregated and analysed. Instead, the sensors can process and aggregate local information and perform some analysis directly within the network. This has great advantages: the local measurements can be verified with respect to outliers, thus only valid information has to be transmitted and centrally processed. There is a series of algorithms, which can be completely processed in the network, such as calculating mean temperature, delineating the boundary of an area or calculating its area (Duckham 2012). In this way, the decentralised sensor network can generate a global perception while only performing local operations. In essence, today's geosensor networks vary in the degree by which neighbourhood information is exploited before the information is communicated to the central server: the extremes are defined by "no local communication", which corresponds to traditional networks, where all data are centrally processed, to "only local communication", where the computation is done in the network itself, requiring—in the worst case—communication between all sensors. It can be expected that future sensor networks will exploit local information to a certain degree, especially, when the amount of data is prohibitively large (e.g. images or point clouds).

In recent years, the OGC has developed specifications for the discovery of and access to sensor data: the SOS (Sensor Observation Service) defined in the Sensor Web Enablement (OGC SWE 2.0) specification (Bröring et al. 2011). This leads to a syntactic interoperability. In order to go beyond that and also allow semantic interoperability, the meaning of the sensors and the measured values has to be made explicit as well (Henson et al. 2009), leading to a semantic sensor web.

In this way, geosensor networks are a new source of spatial information exploitable for different purposes. Even if structural and also semantic annotations are possible through open standards, still, however, variations in accuracy, semantic richness, and redundancy are possible. This places additional demands in terms of interpretation and poses an integration and visualisation problem.

5.2.2 Overview on Aspects and Characteristics of VGI

Through VGI a new class of spatial data is available, which poses new opportunities, but also new challenges for cartography and GIScience. The wealth and diversity of current VGI data are presented in Case study I. We now present a brief description of the essential characteristics of VGI data.

The characteristics are somewhat different to those of authoritative data, such as topographic data or navigation data sets provided by commercial players.

1. **Thematic content/usage spectrum**: Authoritative data sets usually are collected based on a given feature catalogue, which describes in detail the attributes and characteristics of the features contained in the data set and thus provides a kind of ontology for the data. In contrast, most VGI data do not rely on a standardised vocabulary; instead, mostly free text is used. In OSM the feature catalogue is developed via a consensus based discussion. It is worth noting however, that there is a broad spectrum of VGI data available (Case study I), which leads to an unprecedented diversity and semantic richness of spatial data.
2. **Availability, coverage, and homogeneity**: Whereas authoritative data are acquired with the goal of completely covering an area of interest—in the case of topographic maps it is the whole country—this is not the case with VGI data. This is also different from a sensor network, which is usually designed to yield an optimal configuration of observations for the phenomenon of interest. In VGI, the data acquisition depends on the availability of volunteers, therefore the density, depth and semantic richness of the data acquired cannot be

guaranteed. This results in a substantial inhomogeneity. Evaluations concerning the quality of VGI data have been conducted especially w.r.t. OSM data (Haklay 2010; Mondzech and Sester 2011), revealing the dependency of the coverage on different factors, such as country (European users are very active as opposed to US Americans), ease of access to basic information, and availability of freely available, alternative data.

3. **Timeliness of data**: Authoritative and commercial data sets are updated in fixed cycles, depending on the usage of the data. For example, in topographic maps in Germany, the road network is updated every 3 months, whereas the other features only every 3 years. Potentially, user-generated information is of high recency, as users are quick to correct data which is no longer consistent with reality (or else they notify of a bug in the data set[1]). In this way, the community constantly takes care to correct and improve the information. The high timeliness of the data can be exploited for time-critical information such as earthquake-related information which can be gained from Tweets, for example.
4. **Scale range**: VGI data are often large-scale information: this is true for POIs or Tweets. Also OSM data acquired with GPS sensors is of high resolution. In contrast, authoritative data sets are available in different scales—often acquired separately, in recent years also in a (semi-) automatic way via model generalisation. A more detailed discussion on the derivation of different scales in OSM data is given in Case study II.
5. **Quality, reliability, trust, liability**: Authoritative data providers have set up a detailed quality assurance scheme to make sure that their data conforms to given specifications. Similarly, geosensor networks consist of calibrated sensors yielding information complying with standards (Poser and Dransch 2010). In VGI data, this quality assurance process is performed by the users themselves, who notice errors in the data and are able to immediately correct them. Furthermore, there are mechanisms that guard against fraud by excluding users, who have intentionally included errors in the data sets. A critical issue remains one of liability over the data (Rak et al. 2012).
6. **Redundancy**: Some types of VGI do not explicitly seek to provide place information, but more place-related information (e.g. Twitter as opposed to OSM). This will—by nature—lead to redundant data (e.g. ratings of restaurants). Redundancy can also be produced by people who upload their GPS traces.
7. **User-centric information**: Some types of VGI provide highly personal, user-centric and even emotional information, e.g. Flickr images or Tweets. This opens the way to automatically extracting opinions or emotions from the data (Gartner 2012; Tauscher and Neumann 2012).
8. **Spatial reference and geometric representation**: Often VGI data is given in terms of point features with geographical coordinates or addresses, e.g. POIs or Flickr Images. In order to fit those features into the right spatial context, they

[1] www.maps4debugs.openstreetmap.de

have to be visualised on top of a base map, e.g. an authoritative map or with internet maps as background. In this process, the correct topological relationship between the features has to be assured.

These characteristics are relevant for the usage and especially for the visualisation of such data. Mechanisms are needed to enrich it with semantics and to embed it into a spatial context in order to make sense of the data. It is also necessary to filter out irrelevant and incorrect data and to automatically identify the quality of the data. This can be achieved by exploiting the usually high redundancy in the data.

The main challenges are therefore the handling of large amounts of data by providing adequate mechanisms for data generalisation, the extraction of implicit information, and the integration of the data with other, georeferenced, information.

5.3 Aspects of Data Integration

In general, data integration is needed to bring together complementary data sets referring to the same spatial extent, often of the same physical objects, acquired by different organisations, at different times, with different underlying data models and quality levels. The simplest way of data integration is a simple overlay. This is adequate, as long as the geometry of the involved data sets (perfectly) fits, and no geometric and topological errors are introduced through the integration process. Otherwise, matching approaches have to be used to identify corresponding objects and mutually adjust them. Integrating data has several benefits: first, a richer information base is generated, as information (e.g. attributes) from both data sets can be integrated; second, it can also lead to a more consistent data set with higher quality, as through integration possible errors can be detected and eliminated.

The problem of data integration consists of several subproblems: objects from different sources can be integrated, when they match at both a semantic and a geometric level. The first task is to identify the semantic correspondences between object classes. For this, either existing ontologies or object descriptions are used, or the semantic correspondence is inferred from the data themselves (e.g. Noy 2004; Volz 2005; Kieler et al. 2009). Then, features have to correspond with respect to geometric location and geometric properties (e.g. Yuan and Tao 1999), as well as topological characteristics (e.g. Walter and Fritsch 1999; Siriba et al. 2012).

After matching of individual object instances is achieved, different processes can follow from this:

- Information transfer: the attributes of the data sets can be mutually exchanged;
- Data fusion and harmonisation: an integrated geometry of objects can be calculated, possibly leading to a better quality of the averaged geometry. Furthermore, a general transformation of the whole data set can be conducted, and thus also other features, which are not present in the other data set can be

integrated based on their spatial context using rubber-sheeting and/or least-squares adjustment (Safra and Doytsher 2006; Warneke et al. 2011; Dalyot et al. 2012).

Whereas the typical matching cases treat objects of similar geometry (e.g. lines, polygons), in VGI integration often different geometric feature types have to be matched (e.g. matching points with lines, lines with areas, points with areas). This requires additional explicit specifications of constraints to be preserved in the matching process (e.g. the fact that a point has to be on a line, on the first 100 m of the line, or within a specified area; (also discussed in Chap. 3). Furthermore, individual information chunks might not make sense, only their aggregation with similar features may lead to a meaningful concept (e.g. many points at a certain location indicating the importance of a place). Thus, the heterogeneity of VGI data with respect to theme, scale and geometry, leads to the requirement that integration has to take matching, generalisation and interpretation issues into account. The specific challenges are as follows:

- Redundant information: e.g. many traces of travellers indicating traffic infrastructure. Here a statistical integration is required.
- Matching may only be possible at higher levels of aggregation, thus aggregation and/or interpretation of features is required.
- Geometric resolution of the data might differ; therefore, multiscale matching is required.
- Semantic, structural, geometric and topological characteristics have to be preserved while integrating data—with the special constraint, that these characteristics might not be known explicitly.

These aspects will be described in more detail in the following sections. Figure 5.1 shows different levels of correspondences between different representations of data: individual VGI data elements can be recorded multiple times; for example several lines (trajectories) may have been acquired which correspond to one individual river object. Similarly, it might be possible to extract information chunks containing “city centre” from different data sets such as POIs or Tweets. On an individual feature level, the data might not be accurate or reliable enough, however, when the features are confirmed by many other (redundant) measurements, this information can be aggregated and transformed to a new level.[2] In the second level of Fig. 5.1, the hull around the point features delineates the city centre, whereas the aggregated hull around the individual trajectories represents the areal river feature. At this level, the data can be matched to a large-scale topographic data set; it can also be further generalised, for example by using the collapse operation to derive an even smaller scale representation, which can be matched to a small-scale topographic map.

[2] Note that the notion of aggregation is used here to describe the grouping of similar instances of the same feature class to a new, combined representative object.

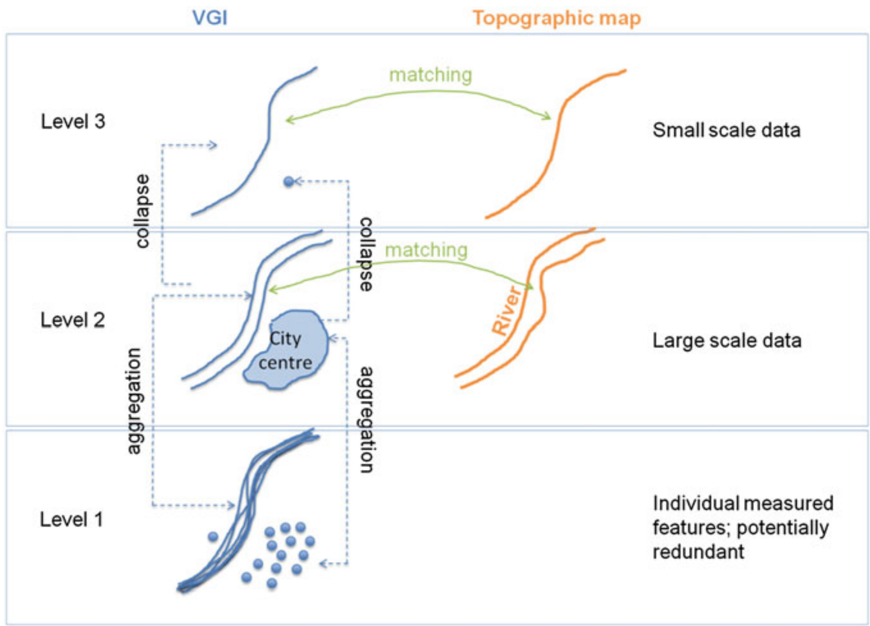

Fig. 5.1 Generalisation and matching of VGI and map data: levels correspond to different representations of the data; some of them can be linked to different scales

5.3.1 Aggregation of VGI

In recent years many researchers have analysed data collected by volunteers such as GPS-traces of hikers or car-drivers, which in principle can be considered as individual digitisations of roads or footpaths. The aggregation of GPS tracks mainly has to deal with the high degree of noise resulting from the low quality of the GPS measurements. This may make it difficult to separate nearby roads and thus reconstruct the underlying structure in the road geometry, e.g. the number of lanes, road width. In order to derive an integrated geometry from the collection of given tracks, aspects of reliability and trust (Sayda 2005) as well as geometric accuracy have to be taken into account (Zhang and Sester 2010).

In the case of road tracks, the goal is to reconstruct the centreline as well as the number of lanes from the noisy road data. Some approaches use histograms in profiles that run orthogonal to the hypothesised road. The mean of the intersection points of the profile with the traces delivers points of the centreline of the road. In order to separate different lanes, Schroedl et al. (2004) proposed a method that finds clusters in specific instances, corresponding to the typical width of lanes. Cao and Krumm (2009) use an approach based on a force model which optimises the displacement of individual tracks towards a modelled centre line. Chen and Krumm (2010) use a Gaussian mixture model (GMM) to model the distribution of

GPS traces across multiple lanes; prior information about lane width and corresponding uncertainty is introduced. Zhang et al. (2012) use a probabilistic relaxation for integrating crowd sourced data. Haunert and Budig (2012) extend an approach of Newson and Krumm (2009) using Markov map matching by also allowing missing road segments to be added (see also Case study III).

Davies et al. (2006) employ a raster-based approach—similar to occupancy grids used in robotics—in order to determine the geometry of roads. In their approach, they also include a temporal component by visualising the fading of roads that are not regularly frequented. In this way, abandoned roads can be identified. Thus they are able to describe the temporal change of objects.

Aggregation of VGI information has been proposed to delineate vernacular areas, i.e. regions that are well known to local people, but often are not represented on the map. They are often described by colloquial names, or may describe areas with fuzzy boundaries, such as the “Schwarzwald” or the “Midlands” (Jones et al. 2008). A popular approach is to use footprints of articles from search engines, Wikipedia articles or Flickr images to generate a collection of points. A kernel density estimation overlaying the points leads to a delineation of the fuzzy regions (Hollenstein and Purves 2010; Mackaness and Chaudhry 2013). In this way, Lüscher and Weibel (2013) successfully delineate city centres. Dahinden and Sester (2009) use the approach to determine the importance of geographic features in order to use it for cartographic visualisation and generalisation. The underlying idea is that the relative frequencies with which the geographic features are mentioned in public knowledge repositories gives an indication of their relative importance.

5.3.2 *Integration of VGI and Topographic Data*

VGI data are often associated with point locations, e.g. Flickr images, Tweets, POIs. In order to provide context for this kind of information, mostly background information in terms of topographic maps are used (as in Case study III). Thus, a challenge is to integrate VGI and the underlying reference data. To this end, ways of matching data are needed, which take known constraints and semantics into account. For example, Stigmar (2005) describes how topographic information (which was in a non-routable form) could be integrated with a routable data set. This was achieved using matching techniques, where the closest (and topologically best fitting) routes have been assigned to each other. In the final stage, only the routing instruction was transferred as an attribute—and not the geometry.

Another example is the semantic enrichment of 3D city models proposed by Smart et al. (2011). Geometrically exact ground plans from cadastres were augmented by information from Wikipedia to provide a semantically richer data set. The challenge in the approach was the matching of the VGI point information with polygons from the cadastre. To this end, the authors proposed a fuzzy matching approach taking the distance of the points to the corresponding polygons into account. Additional criteria, such as importance values derived on the one hand from the complexity of the ground plan (assuming that important or interesting objects such as a castle often have a complex ground plan) were also taken into account. The relative importance can also be indicated by the fact that several VGI sources relate to an object. In this way, implicit constraints can be stored. For example the fact that VGI annotations or POIs are mostly placed inside an areal object or at least in its vicinity is coded and resolved in a fuzzy matching approach.

Finally, Jaara et al. (2011) present a concept for integrating thematic data layers with background topographic information. They point out that constraints have to be preserved or even exaggerated. For example it is important to retain the position of a POI to the left or right of a road, or its location with respect to an inflection point of the road (see also Chap. 3).

5.3.3 Matching Multiscale Data

The integrated treatment of data from multiple scales has been investigated mostly in the context of topographic maps of different scales. One important research issue in the recent years was the creation and exploitation of MRDB—multiple resolution/representation databases. Different approaches have been proposed to generate the links between objects of different scales, e.g. by Devogele et al. (1996), Weibel and Dutton (2005), Hampe et al. (2004), Mustière and Devogele (2008), Shereen et al. (2009). The challenge is to take different geometric representations of the objects into account. For example, Kieler et al. (2009) implemented an algorithm that collapses rivers represented as polygons before matching them to the small-scale line objects in another data set. Similarly, different cardinalities in matching have to be taken into account, as typically several objects are aggregated to represent an object in the smaller scale, thus leading to an n:1 or n:m-matching situation.

5.4 The Visualisation and Generalisation of VGI

5.4.1 Visualisation of VGI

Visualisation of VGI data, possibly together with underlying base data, poses several requirements, which are substantially different from visualising conventional maps. The first issue considers the graphical design and adequate information selection, the second issue concerns the communication of the inhomogeneous quality of that information.

5.4.1.1 Graphical Design, Information Visualisation

In contrast to topographic maps, often no commonly accepted feature catalogue is available, nor an agreed signature catalogue or visualisation and generalisation rules, which describe how VGI should be graphically presented.

Another issue is the often heterogeneous information density of VGI, e.g. when visualising PoIs: the PoIs are often clustered in the city centre, whereas typically there are no or only a few features in surrounding regions. This requires an adequate information selection and placement. Different options have been proposed, for example by Slingsby et al. (2007). Other approaches try to use visualisations such as tag clouds to present the available information at a certain location (Paelke et al. 2012). These visualisations are also scale dependent and have to change as the user zooms in and out. Possible mechanisms for selection of adequate information and symbol sizes have been proposed by Bereuter et al. (2012) using a quadtree principle. Other approaches use focus maps to arrange (and aggregate) PoI-labels (Fink et al. 2012).

5.4.1.2 Visualisation of Data Quality

Trying to visualise the heterogeneity inherent in VGI data is a big challenge. A fundamental cartographic principle is to provide a homogeneous view of the real situation, which implies that a blank area in the map indicates the absence of information. This can of course not be guaranteed in VGI, as it can also indicate an area which just has not yet been mapped. Thus the challenge is firstly to identify where there are gaps in the data, i.e. areas, which have not yet been mapped. One option to detect these areas is to analyse the density of data (possibly in relation to the underlying topography) and thus gain a sense of areas that are unlikely to have been mapped. A second issue is to determine the level of detail of a certain piece of information. In other words, does a Wikipedia article or a photo refer to a small reference object or to a larger area? Here, references to gazetteers may help resolve this issue.

Once this meta-information about data quality is available, different techniques can be used to graphically present this information (MacEachren 1992). Senaratne and Gerharz (2011) describe an evaluation of different possibilities to visualise uncertainty or different qualities in the data. Kinkeldey and Schiewe (2012) use so-called noise annotation lines to create an overlay grid indicating the quality of the underlying data.

5.4.2 Data Generalisation and Adaptation of Scale

In order to integrate data, possibly different scales of base data and VGI data have to be taken into account. Different cases can occur: either the VGI data are more detailed or the base data are more detailed. In addition, the level of detail in the VGI data can vary from region to region, thus there is no homogeneous scale.

Broadly, generalisation is needed for several reasons (Jaara et al. 2011):

- In order to increase the consistency by changing the scale or the level of detail of layers;
- in order to decrease the level of detail in one or more layers, to adapt it to the context of use, the intended display scale and within the associated graphical constraints. Case study III presents an approach for the integration of VGI which has the potential of automatically adapting to the scale of the base map;
- generalisation is needed to visualise data at a smaller scale; this generalisation process has to take the characteristics of the individual objects, but also their relations into account by preserving or even enhancing them. Case study II describes the challenges of generalising VGI data—something that is quite different from authoritative data.

In order to create a consistent visualisation, we have the following options:

- creation of one homogeneous scale via generalisation;
- communication of difference in scales so that for the reader it is clear for which generalisation level the information is intended;
- creation of a 'focus lens', i.e. keeping both scales, but indicating, where which scale is used with the help of adequate graphical variables.

The main generalisation operations needed are as follows (the order indicates their importance and use in current approaches):

- **Selection and typification**: the most important operation, as it has the potential to significantly reduce the amount of data;
- **Aggregation/grouping**: this operation can be applied to generate a consistent object geometry by averaging several geometries; it can also be used to classify and aggregate objects of similar categories;
- **Dimensional change**: is an operation, which is applied in different ways the first is the grouping of several PoIs to form an area (e.g. the city centre). This can be considered as a change from point information to areal information. The second and more typical change is from higher to lower dimensions e.g. representing an areal market place as a point at a lower scale.
- **Displacement**: is needed in order to eliminate spatial conflicts between objects.
- **Enhancement**: can be applied to visually attract the attention to important, but small objects.

Case study II describes the use of generalisation operations in the case of OSM data.

5.5 Case Study I: VGI Platforms and Data Generalisation

Jamal Jokar Arsanjani

Undoubtedly, Web 2.0 technologies have provided a tremendous amount of user-generated content, which accounts for an interactive and alternative source of information from and for, the public. Among the contributed information, a large amount of information contains geolocated information or information about geographical objects. We are facing a new era in gathering information about objects and events disseminated via the web. Thanks to online mapping projects, individuals are able to digitise objects of interest from base maps via web mapping services (WMS) or alternatively collect information about geographical entities via GPS-enabled devices and then share them with the public. The term VGI was coined by Goodchild (2007). In the case where the individuals' contributions are taken without their awareness, this sort of information is called contributed geographic information (CGI). Up till now, citizens science projects, public participatory GIS projects (PPGIS: Carver 2001; Rambaldi et al. 2004), collaborative mapping projects (CMPs: Rouse et al. 2007), and crowdsourcing (Heipke 2010) have been the main sources of VGI and CGI. Although the new given possibilities by VGI, it does not offer the collected data via classical, consistent, and standardised data structure. Some examples such as georeferenced photographs (Flickr, Panoramio), videos, check-ins, and tweets can be mentioned here.

A large number of VGI-based services have been launched for covering a variety of disciplines—taking different forms of data from individuals and sharing them in different ways. It is important to understand the limits of VGI, and identify new research challenges. This Case study is intended to identify some representative VGI services and to consider their individual functionalities, fitness-for-use, and identify which types of information are provided by them. The accessible data from VGI will be considered from a cartographic perspective (e.g. generalisation and visualisation). The VGI services are categorised according to several classification systems based on their usage.

5.5.1 Characterisation and Categorisation of Popular VGI Services

Each individual VGI service collects a specific source of data and shares the collected information through a particular data type. For instance, OSM, Wikimapia, and Google Map Maker provide an opportunity to map the world into different feature classes, which are represented as point, line, polygon data types. Flickr and Panoramio provide geolocated photos along with their location and attribute tags. A set of VGI services is summarised below.

5.5.1.1 World Mapping Projects

OpenStreetMap is a collaborative mapping project (CMP), which seeks to create a free editable map of the world. It has attracted more than 1.3 million users so far. Users are able to digitise the geographical objects from WMS into points, polylines, and polygons in addition to importing GPS tracks recorded by smart devices and shapefiles from official sources (e.g. CORINE data in France and TIGER data in the USA). The features are structured as POIs, places, roads, railways, waterways, natural, land use, and buildings (http://wiki.openstreetmap.org/wiki/Map_Features). The compressed and uncompressed versions of OSM data can be downloaded via either OSM-API, or querying data through a plug-in in GIS softwares, or through Geofabrik and CloudMade websites (Ramm et al. 2010). OSM features have been further implemented in a variety of applications such as routing (www.Openrouteservice.org), 3D modelling (www.OSM-3d.org), disaster management (Yates and Paquette 2011), and land use mapping (Jokar Arsanjani et al. 2013b). OSM has been extended to include the mapping of the maritime—via *OpenSeaMap*.

Wikimapia is another popular CMP, which enables the marking of geographical objects and describing them in polygonal form. The project's motto is *Let's describe the whole world!* and has over 1.8 million users. Wikimapia API and (Motomapia.com) enable users to record the location, time stamp, category, tags, and descriptions of objects. *Google Map Maker* is also a CMP sponsored by Google to promote the quality of Google maps. Like OSM, users create and edit the objects into points, polylines, and polygons. It has two major differences compared with the above-mentioned services:

(a) it doesn't cover all regions in the world and only certain countries are covered (https://services.google.com/fb/forms/mapmakerdatadownload/);
(b) the contributed data are not downloadable.

5.5.1.2 Social Media Mapping

Social media mapping services allow users to share text-based messages, postings, photos, and videos captured by GPS-enabled devices (e.g. digital cameras and smart phones). *Flickr* helps people to share their photos with the public and has more than 51 million registered users and 6 billion uploaded photos, of which more than 200 million are geotagged. The coordinates are in point form (latitudes and longitudes). Via Flickr API, coordinates, time stamps, tags, and textual descriptions of the photos can be recorded. *Panoramio* is also a photo sharing platform which contains geolocated photos. Like Flickr, it provides tags assigned to photos and, the photos can be downloaded through its API.

Twitter is a platform which enables people to share their tweets, expressing ideas, reports, news, and events around the world. At present there are over 500 million registered users and over 300 million tweets are generated daily.

Geolocated tweets can be mapped as point features together with textual attributes. In addition to the coordinates and textual content of tweets, time stamps of the tweets and user profiles are available via the Twitter API. Twitter has frequently been leveraged in event detection e.g. earthquake, crowd behaviour, political campaigns, and disaster management.

Facebook is the most popular social network, which is dedicated to helping people connect and share their interests and information. Currently it has over one billion users. In addition to posts, geolocated photos and videos are available, which can be mapped as point features. The Facebook API provides the content, location, time stamp, user profile of the posts, photos and short videos. *Foursquare* is another popular social network that enables users to check-in at venues and share check-ins. Each check-in has coordinates, a time stamp, and attributes that can be accessed as point features via the Foursquare API. Possibilities and challenges of exploiting social networks are discussed by Roick and Heuser (2013).

YouTube is the most popular video-sharing website, which enables users to upload videos and edit videos by adding tags and descriptions. Similar to photos in Flickr, the videos themselves are uploaded to YouTube along with their metadata. Such metadata which contain time stamps, tags, coordinates, user names and textual descriptions are acquirable via the YouTube API as point features.

5.5.1.3 Environmental and Ecological Monitoring

Eye on Earth is a large and ambitious project supported by the European Environment Agency and dedicated to building a crowdsourced map for environmental monitoring. Volunteers make contributions by attributing their feelings on the quality of air, water, noise and nature. Observation sites marked on the base map of Eye on Earth are recorded as point features enabling various quality maps to be generated.

Geo-Wiki has been launched by the International Institute for Applied Systems Analysis (IIASA) in order to improve the accuracy of global land cover maps through volunteers (Fritz et al. 2012), who are asked to add their information on habitat and ecosystems. Volunteers validate the land cover maps by comparing global land cover maps with Google Earth together with their own knowledge. The data are recorded as point features and can be downloaded directly from the website.

eBird,Breeding Bird Surveys,Christmas Bird Count, and *Wildfinder* are free online databases of bird observations, recorded as point features and offering real-time data about bird distributions collected by citizens. Such data facilitates biological research (Sullivan et al. 2009; Scofield et al. 2012). The data are downloadable directly from the respective websites.

Tracks4Africa intends to map African ecosystem reliably from experienced and responsible eco-travellers who are based in the field. The main source is via GPS mapping of eco-destinations in rural and remote Africa of outdoor activities e.g.

hiking, mountaineering, river rafting, scuba diving, bird watching, paragliding, and green land travel through points of interest and thematic maps.

5.5.1.4 Weather Mapping

Weather mapping projects are platforms that enable users to provide real time descriptions of weather conditions as point records. The most practical weather mapping project is called 360.*org* which seeks to harvest weather information from weather stations, universities, and amateurs in order to broadcast weather predictions to the public. The website allows users to download the data directly.

5.5.1.5 Crisis and Disaster Mapping

Ushahidi is an application that collects reports on different crises from volunteers. Originally it was used to visualise incidents of violence and peace efforts throughout Kenya from 45,000 users, and now supports several events around the world, which can be downloaded from the Ushahidi website. An evident example of its effectiveness was proven in the Haiti earthquake (Zook et al. 2010; Yates and Paquette 2011). *Crowdmap* is also a platform used to collect crowd-sourced crisis information. *Did You Feel It?* is built to encourage people who actually experience earthquakes to estimate the effects and extent of damage of earthquakes. An intensity map is generated by combining the information provided by both official agencies and internet users. The geographic features are contributed as points and polygons representing the locations and intensities of affected regions (Atkinson and Wald 2007).

5.5.1.6 Crime Mapping and Tracking

WikiCrimes offers users interactive maps to anonymously report crimes and pinpoint their position in order to identify crime hotspots. Users are predominantly from Latin America. The location, the type and density of crime are accessible as point features. The data of WikiCrimes is downloadable from its website.

5.5.1.7 Outdoor Activity Mapping

Outdoor activity mapping projects allow citizens to share GPS tracks of their leisure trips and recreation spots such as running, walking, hiking, and biking. *Wikiloc* with more than 700,000 users is very popular for discovering and sharing the best trails for outdoor activities, which additionally includes PoIs, elevation profile, distances, accumulated altitude, and images. The information is recorded as polylines and points. Similar projects such as *EveryTrail, MapMyRun, Endomond,* and *Map My Tracks* have similar functionalities.

US Fish Finder allows users to add photos and fishing hot spots which are accessible to the public. The application provides a variety of information for people such as insight into fishing hotspots, plotting lakes, rivers and streams and directions to them, access and boat launches, uploading photos of specific locations, topographic maps, and tide information via point and polyline features.

5.5.1.8 Business Mapping

Yelp is a local directory service with social networking and user reviews which helps users to find local businesses such as restaurants, bars, hotels, and petrol stations. Similar to the venues in Foursquare, business venues in Yelp receive reviews, ratings and comments from users. Geolocated business venues are considered as geographic points. Via the Yelp API, the attributes such as business name, rating, review count, address, phone number, neighbourhood, category, latitude and longitude, and reviews are available. Similarly, *Where* is used by local businesses to reach local audiences by providing data of various features. It also enables check-ins on smart phones again considered as geographic points.

5.5.1.9 Transportation Mapping

TomTom Map Share enables users to avoid congested sections of roads. It allows users to report changes and to share them with others via their mobile devices or via an online Map Share reporter tool offered by TomTom. The reported information on road changes includes modifications of speed limits, new street names, blocked roads, new traffic directions and altered turn restrictions. The geolocated roads are considered as polylines. TomTom makes such information changes visible but not downloadable to the public. Similar examples include *TeleAtlas map insight* and *Navteq map reporters* (Coleman et al. 2009). *Here.com* also allows the public to edit Navteq and Nokia maps. *Waze* is a community-based traffic and navigation application which enables users to share real-time traffic and road information. The information includes the users' historical routings, real-time traffic jams, and petrol prices. *Trapster* provides the opportunity for people to report speed traps, red lights and speed cameras, accidents, and road hazards. It has a user base of nearly 18 million users. The geolocated traffic traps are mapped as geographic points.

Table 5.1 gives a summary of the dominant feature types acquired by the different VGI services. Predominantly the information is collected as point

Table 5.1 Dominant geometric features types acquired in different VGI services

Data source	Points	Polylines	Polygons
World mapping projects	X	X	X
Social media mapping	X	X	
Environmental mapping	X		X
Weather mapping	X		
Crisis and disaster mapping	X		X
Crime mapping	X		
Outdoor activity mapping	X	X	X
Business mapping	X		
Transportation mapping	X	X	

information. This characteristic is relevant for subsequent visualisation and generalisation procedures.

5.5.2 *Data Quality Aspects*

Accuracy indicates by how much the data on a map or in a geodatabase matches with the reference data/values, whilst *precision* notes the level of correctness of information in a database (van Oort 2006; Geravis et al. 2009). However, the concept of "data quality" is sometimes inappropriately interpreted as data precision, uncertainty, or error. Although geodata with high locational precision are often called high-quality data, the sense of quality is beyond just the concept of locational precision (van Oort 2006).

The quality of geodata should be internally and externally taken into account. Internal quality takes the data production standards and specifications into consideration by detecting the errors in the data. Some standard organisations such as ISO, ICA, FGDC, and CEN define internal quality based on five mutual aspects (a) attribute accuracy, (b) positional accuracy, (c) temporal accuracy, (d) logical consistency, and (e) completeness (Guptill and Morrison 1995). These data properties are usually given to the users through metadata files attached to datasets by the producers (Devillers et al. 2007).

External quality considers whether a dataset is suitable enough for a specific purpose—termed "Fitness for Use". In practice a dataset should meet or exceed some expectations from end users regardless each individual internal data quality aspect (Devillers and Jeansoulin 2006). Quantitative VGI data quality analyses have been applied mainly to OpenStreetMap e.g. Haklay (2010), Helbich et al. (2012), Jokar Arsanjani et al. (2013a) in comparison with proprietary datasets. However, Goodchild and Li (2012) propose using VGI as wisdom of crowd to evaluate the quality of VGI.

5.5.3 *Challenges of Integration, Generalisation, and Visualisation of VGI*

The possible challenges of integration, generalisation, and visualisation of VGI are:

(a) Integration: Data integration is intended to integrate data from different sources in order to provide a consistent database of objects. This concept should be considered differently in the case of VGI, because several users may separately report a single object to the same VGI platform with slightly different positional placement and therefore different location and attribute information. It might become even more confusing when different VGI sources exist and dissimilar versions of objects and their attributes are collected. The shared information might not necessarily match at some points and thus data quality analysis must be used to filter noisy, redundant, and inaccurate data (Yuan and Tao 1999). Applying majority filters enables us to resolve redundant attributes and data generalisation and matching resolves the positional redundancy of the objects (Fig. 5.2). Despite the difficulties of gathering the correct information from VGI, it certainly helps (a) to improve the positional accuracy of the objects or to record the latest positional shifts, and (b) to enrich the attributes and metadata of the datasets. Having more reports on the same objects might suggest new indicative concepts such as place popularity, landmarks, events, trajectories, users' interests and behaviours.

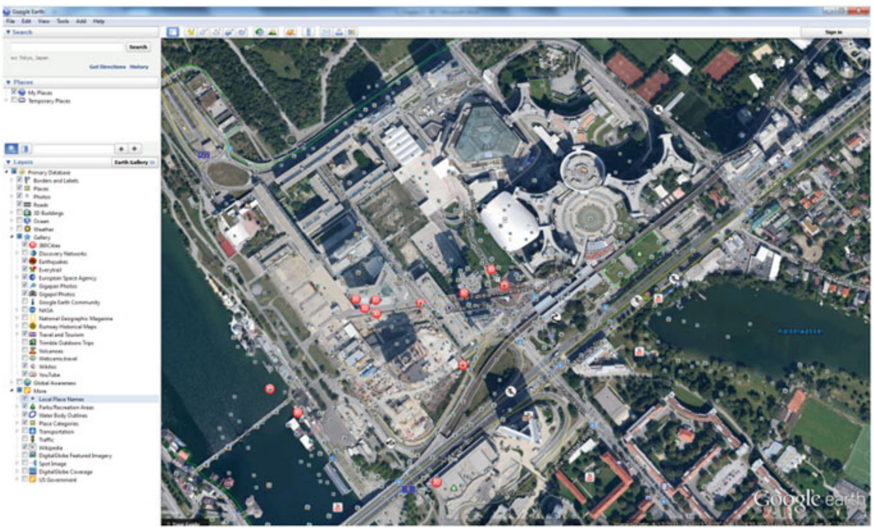

Fig. 5.2 Concurrent visualisation of some VGI data via Google earth

(b) Generalisation: The intended level of detail (LoD) can be an indicator for identifying the degree of generalisation required (Mackaness et al. 2007). Contributions to VGI come from diverse and heterogeneous sources (GPS tracks, online digitisation, shapefiles import) and therefore, each contribution has its own LoD. For instance, GPS tracks are captured from different devices at different precision and accuracy criteria and diverse signal receiving conditions. Contributions may have been digitised at dissimilar zoom levels. In other words, one side of a linear object could have dense nodes within a certain distance, which fits into a data scale 1:A, and the opposite side has widely separated nodes, which fits into data scale 1:B. This results in having features, which need generalisation techniques to make them homogeneous. Each shapefile has a certain scale and online digitisation is possible at certain levels of zoom. For example OSM has more than 10 (http://wiki.openstreetmap.org/wiki/Zoom_levels). The WMS contain different types of aerial and satellite images at different spatial and temporal resolutions. Therefore, the VGI data are heterogeneous in terms of source, producers' expertise, accuracy, and LoD (Girres and Touya 2010; Haklay 2010). This suggests that VGI data do not fall into certain scale ranges and accordingly cartographic generalisation might not be consistently applied. Since VGI might offer richer datasets than classical datasets in some cases, the aforementioned limitations must be tackled by developing new techniques such as identifying features with different accuracy, shape, and spatial relations of features in order to assign the most relevant scale and also switch between different scales and times tamps via multiple representation databases (MRDB: Balley et al. 2004; Hahmann and Burghardt 2010). This helps to harmonise the LoD of mapped features to facilitate the generalisation process.

(c) Visualisation: For VGI, the classical visualisation techniques need to be advanced in order to address these variabilities. Scale-dependent visualisation techniques enable us to represent objects at different scales and time stamps. We also need ways of visualising tweets that express events and trajectories as well as Flickr photos tags. Since most of VGI services provide their data through PoIs, the latent information must be extracted from the PoIs through point pattern analysis (Sengstock and Gertz 2012). Some VGI websites act as a pool in which data can be both collected and visualised. Google Earth is an example of this (Fig. 5.2).

5.5.4 Summary

VGI provides information from people on diverse objects, events, and activities. VGI is offering free, new, and never-collected information as well as new data types and structures that never existed before. VGI is a data pool with heterogeneous sources, heterogeneous data structures, heterogeneous producers, and heterogeneous quality. Therefore, it asks for new cartographical techniques on data integration, data fusion, data generalisation, and data visualisation. It is important to note that VGI is provided by people with varying expertise which raises some concerns on the credibility and reliability of VGI (Jokar Arsanjani et al. 2013a).

New techniques and algorithms on assessing its quality and fitness for use have been increasingly developed and continue to be improved. As more documents, practices, and researches on VGI are released and published, more advancements and developments of algorithms and approaches are expected.

5.6 Case Study II: Generalisation within the OpenStreetMap Project Compared to the Generalisation of Authoritative Data

Ralf Klammer and Dirk Burghardt

This Case study will provide a structured comparison between the application of generalisation strategies within the crowd controlled structure of the OpenStreetMap project to the officially controlled structure of NMAs. This will be done by illustrating the essentially applied generalisation processes of OpenStreetMap in contrast to the appropriate analogies of the conceptual model of digital landscape and cartographic models introduced by Grünreich (1992). Beginning with an outline of basic principles about both systems this Case study will compare the structure of OpenStreetMap and NMAs in the second section and illustrate the application of automated generalisation within the OpenStreetMap project in the third major section.

5.6.1 Basics Principles of Authoritative versus VGI Data

The basic concept of OpenStreetMap (OSM) is the supply of a freely available world wide spatial dataset by activating nonprofessional, volunteered, 'hobbyist-surveyors' to collect spatial information of their immediate environment using simple GPS devices or to convey spatial information of personally inaccessible regions using aerial images. The result is a continuously updated central database at global coverage. The quality and quantity of this free available dataset is mainly characterised by heterogeneous spatial density (e.g. *industrialised versus developing countries*), geometrical quality (e.g. *urban versus rural areas*) (Girres and Touya 2010; Neis et al. 2012) or feature type assignment (e.g. *local versus global tagging of administrative boundaries*[3]) (Codescu et al. 2011). Beyond that, OSM is

[3] http://wiki.openstreetmap.org/wiki/Tag:boundary%3Dadministrative#10_admin_level_values_for_specific_countries

a web based crowdsourcing project, with a huge variety of map applications[4] and utilised rendering software.[5] This diversity precludes the description of target data, maps and courses of action by a finite common generalisation strategy. Accordingly, this Case study will focus on the basic map visualisation of OSM data on openstreetmap.org as well as the most common usages of OSM datasets, described in the corresponding project documentations (e.g. *wikis*[6] *and tutorials*[7]).

The so called Grünreich-Model (Fig. 5.3), introduced in the early 1990s, is a fundamental conceptual model for map generalisation providing a basic overview on the applied generalisation strategy within official cartography. Its differentiation between object-, model- and cartographic generalisation facilitates a fundamental consideration of all generalisation procedures that arise from spatial data processing. It begins with the capturing of spatial data in the field and ends with their information visualisation on a map. This model is still generally accepted today (e.g. Stoter et al. 2010; Bobzien et al. 2007; Chap. 11) and is especially adapted to the applied data management and visualisation of the German authoritative topographic cartographic information system (ATKIS) (Grünreich 1992). Accordingly, it provides a good base for a structured comparison of the application of map generalisation to authoritative and crowd controlled data, as it is more generic than for example the conceptual models of and McMaster and Shea (1992) or Brassel and Weibel (1988).

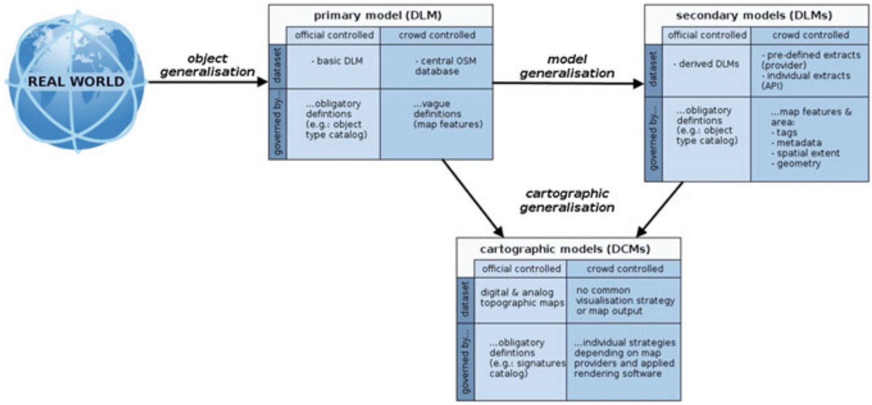

Fig. 5.3 The adapted Grünreich model in practice—NMAs versus OpenStreetMap

4 http://wiki.openstreetmap.org/wiki/Maps

5 http://wiki.openstreetmap.org/wiki/Renderers_feature_list

6 http://wiki.openstreetmap.org/wiki/Main_Page

7 http://www.osmfoundation.org/wiki/Main_Page

5.6.2 Comparison of Influencing Conditions on Generalisation

The basic conditions, that lead to the identification of structural differences in map generalisation defined by official and crowd structures, are contrasted in Fig. 5.3. NMAs traditionally pursue a central organised generalisation strategy with the primary aim of providing analog and digital topographic maps at different levels of detail, while efforts on supplying web-accessible maps are also aspired to (e.g. Sects. 11.6 and 11.7). At the other end of the spectrum, OSM is basically a decentralised crowdsourcing project with individual generalisation strategies, dependent on the data handler. Although a central unit exists in the form of the OpenStreetMap Foundation (OSMF), it does not control the project.[8] The primary map application of OSM data are interactive web maps commonly provided at up to 18 different scales, colloquially called zoom levels, mostly with a global coverage. NMAs provide maps of their country with up to 8 different scales (Chap. 11). The major difference between both is centered around how the data are processed. The basic tile-based map visualisation of OSM on openstreetmap.org is aimed at an immediate (Ramm 2012) visualisation[9] rendered completely automatically using data directly from the central OSM database.[10] The disadvantage of this performance driven rendering strategy is a rudimentary application of fundamental generalisation operators, resulting in restrictions to the graphical quality of maps (Touya 2012). The visualisation strategy of NMAs is generally geared to user oriented requirements (Chap. 2) which is why the whole set of generalisation operators is applied to produce map visualisations tailored to the user. As a consequence, the data processing is done, with exceptions outlined in Sect. 11.7, in a semi-automatic manner with interactive completion of the automated generalisation processes (Table 5.2).

5.6.3 Automated Generalisation in OpenStreetMap

To get a better sense of generalisation within the OSM project, this section will analyse OSM compared to the three basic operations—object, model and cartographic generalisation—of the Grünreich model.

Object generalisation is defined as the decision process by which spatial information has to be gathered, and which offers the highest spatial and semantic resolution (Fig. 5.3). This dataset is termed the primary digital landscape model (DLM). Object generalisation is realised in both the official and crowd controlled

---

[8] See Footnote 7

[9] http://munin.openstreetmap.org/openstreetmap/yevaud.openstreetmap/renderd_queue.html

[10] https://help.openstreetmap.org/questions/178/how-often-does-the-main-mapnik-map-get-updated

Table 5.2 Structural differences between official and crowd controlled map generalisation

	Officially controlled	Crowd controlled
Organisation	Central	Decentralised
Primary map application	Topographic maps (analog and digital)	Web maps (miscellaneous applications)
Scale range	~5–8 different scales	~17–20 different scales
Data coverage	National (regional)	Global
Primal guidance	Graphical quality (legibility)	High performance
Data processing	Partly automated with manually controlled interaction	Completely automated
Average update cycle	Yearly	By the minute
Modelling	MRDB-approaches (primary and derived secondary DLMs)	Rendering from a central database (primary DLM)
Applied operators	Complete set of generalisation operators	Selection, (Re-) classification

systems in a similar way. However, they are very different in detail. In officially controlled structures these decisions are determined by obligatory rules (e.g. *object type catalogs*) established within working groups, to which professionals must adhere when they gather data for the primary model. In OSM these rules are vaguely specified by OSM map features.[11] These are termed as "tags" and are used to assign features within the internal OSM data structure (Ramm et al. 2010). There is no obligation to respect any of them and each OSM member is generally able to define their own map features independently. Nonetheless, most members are geared to the commonly approved specifications in OSM practice and the introduction of new tags is often elaborately discussed within the OSM community via proposals.[12] The results of the object generalisation are in both cases a primary digital landscape model, while the primary model of OSM—the central OSM database—has a higher semantic resolution (e.g. *defibrillator*[13]). A second major difference is the type of person who collects the data: They vary from skilled and experienced surveyors, to unskilled volunteers with different levels of experience who control and increase the data quality collaboratively (Haklay et al. 2010) by relying on "The Wisdom of the Crowds" (Surowiecki 2004). This significantly influences the homogeneity of the primary DLM of OSM (Siebritz et al. 2012).

Model generalisation is defined as the derivation of secondary DLMs from the primary DLM, with a lower spatial and semantic resolution (Fig. 5.3). This is applied by different strategies in the NMA work practice but can be abstracted to a derivation of generalised databases adjusted to the production of small scale maps (Chap. 11). Accordingly, the model resolution is reduced by semantic (e.g. *feature*

[11] http://wiki.openstreetmap.org/wiki/Map_Features

[12] http://wiki.openstreetmap.org/wiki/Proposed_features

[13] http://wiki.openstreetmap.org/wiki/Tag:medical%3Daed

types) and geometric (e.g. *minimum size*) object selection. Applied generalisation operators aggregate areas, simplify geometries or build new object classes (classification). Many NMAs are already able to implement model generalisation as a fully automated process within their production workflow (e.g. Sect. 11.4). This automation effects the application of generalisation operators additionally, depending on the availability of appropriate algorithms and tools. In addition, storing and managing different forms of object representation implies the need for a specific MRDB concept (Kilpeläinen 1992). This is especially important for the update of secondary DLMs.

A comparable common course of action does not exist for OSM. Certain private initiatives and companies provide derived datasets, respectively secondary DLMs, with a lower spatial and semantic resolution than the central database. Two different types of derivation strategies exist. The first one is the derivation and supply of spatially and semantically filtered data extracts. The currently most common provider of such extracts is the company "Geofabrik",[14] who supplies daily[15] updated and completely automatically derived extracts from the central OSM database. This is especially advantageous for less experienced users of OSM data, as the whole OSM dataset is fragmented to administrative units (country specific datasets). The main disadvantage of these extracts is the lack of influence that a user has on the generalisation constraints as they are defined by the provider.

This limitation is overcome by the second derivation strategy, which provides the download of data extracts via application programming interface (API). The user can define individual data requests, adjust the degree of resolution reduction but needs programming skills and experience in using an API. Although the central database offers an integrated API,[16] the OverpassAPI[17] is currently the most stable and requested for deriving individual OSM data extracts. This implementation offers user defined reclassification of feature classes in addition to pure semantic filtering and uses a Quadtile[18] approach, based on Quadtrees (Samet 2005), for high performance execution of queries. Both strategies offer fully automated implementations of model generalisation with semantic selection—related to feature types or OSM metadata (e.g. *editor, time*)—and spatial selection—related to spatial extents or specific geometries (e.g. *country outline*). Consequently each extract can be termed as secondary OSM-DLM derived from the primary OSM-DLM (central OSM database). A specific MRDB concept for managing different forms of representations is not applied within the OSM project. However, objects can be linked by a unique identifier which is assigned to only one OSM feature and does not change during a features life-cycle. The uniqueness of features is a core concept of OSM as the identifier is used to build complex

14 http://www.geofabrik.de/en/index.html

15 http://blog.geofabrik.de/de/?p=75

16 http://wiki.openstreetmap.org/wiki/API

17 http://overpass-api.de/

18 http://wiki.openstreetmap.org/wiki/QuadTiles

geometries from the specific OSM data format that only consists of nodes, ways and relations (Ramm et al. 2010).

Cartographic generalisation is characterised as the creation of digital cartographic models (DCM) by applying object symbolisation to the features of a specific DLM. Digital landscape models are just the preliminary stage of a final map visualisation where only the resolution of datasets is generalised without considering any legibility issues. Cartographic generalisation has, in contrast, a direct impact on the cartographic quality of a map as it includes the evaluation of feature representation and behaviour in a spatial context. This is done by contrasting object symbolisation with map specifications and user requirements, such as legibility (Chap. 2). This evaluation (Chap. 9) as well as appropriate geometric modifications can be applied generally to individual objects (independent generalisation) and sets of objects (contextual generalisation) implemented as manual, semi- or fully-automatic processes. An automatic computational recognition of graphical deficits needs accurately predefined constraints that characterise graphical object restrictions (e.g. *minimum sizes or distances*). These constraints would be evaluated by a map author visually within a manual evaluation.

Cartographic generalisation is driven by object symbolisation which is true for the official as well as for the crowd controlled structures. NMAs seek completely automatic workflows though the final evaluation and adjustment of cartographic quality is mostly done manually (Sect. 11.4). Completely automatic implementations have to accept a loss of graphical quality to provide a shorter refresh period (Sect. 11.3). Accordingly, the decision of implementing automatic or manual processing is currently a compromise between graphical quality and performance.

This also applies to the OSM project. Manual processes are almost completely ruled out in advance because of the huge amounts of data, in the specific case of the basic OSM map. Automatic evaluation or geometric modification is currently not aspired to in the general strategy or even implemented in the basic OSM map. The actual course of action for rendering a tile-based map, such as the basic OSM map, is as follows: Mapnik,[19] the main library, is applied to render all necessary map tiles, by obtaining OSM data from a regularly updated, locally managed version of the central database. The corresponding object symbolisation is defined within specific style files. Claims on the automatic application of generalisation operators have to be integrated to these definitions. Current implementations only contains fundamental generalisation operators—semantic selection and semantic reclassification—operators which are applied as model generalisation by NMAs. That means, the style definition contains a reference to the whole set of data (primary model), the corresponding data filter (e.g. [*highway*] = ‘*residential*’) and the declaration of the scale range for displaying the filtered features (e.g. *<MaxScaleDenominator>*25,000*</MaxScaleDenominator>*). The map author is responsible for the declaration of these definitions. The rendering tool (Mapnik) fulfills the data filtering. The generalisation process—semantic filtering and

[19] http://mapnik.org/

reclassification—is shifted to the rendering, and applied on-the-fly. Additionally, Mapnik has integrated methods for optimised label placement and simplification of linear geometries, but the OSM community has currently no common strategy on how to utilise these methods.

This course of action does not precisely fit with the definition of cartographic generalisation, as there is no evaluation and modification of graphical deficits that result from object symbolisation. Accordingly, it is only possible to include the described basic tile-based generalisation operations of OSM by weakening the initial definition of cartographic generalisation to a pragmatic definition. As a result, cartographic generalisation has to be characterised as the general application of generalisation operators for the visualisation of spatial information on a map.

5.7 Case Study III: Matching GPS Trajectories with Incomplete User-Generated Road Data

Jan-Henrik Haunert

5.7.1 Motivation

Map matching is the problem of finding a path in a given road data set that corresponds to a given trajectory. The trajectory is a sequence of positions with time stamps, which usually has been recorded with GPS. Several map matching algorithms have been developed that work well, even if the GPS measurements are noisy and the sampling rate of the GPS measurements is low (Newson and Krumm 2009; Quddus et al. 2007). The problem of incomplete or inaccurate road data, however, has seldom been addressed. Further research is needed, in particular to develop map matching algorithms that are suited for user-generated road data, which are often incomplete. In fact, user-generated data are often *more* detailed than commercial or official data and, in many regions, are suited for pedestrian or bicycle navigation. The quality of user-generated data varies, however, depending on the number of contributing users in a particular region, the time they spend on the project, and their skills.

Every user who contributes to a mapping project generalises, simply because he or she has to decide which objects to include in the map and how to represent them. Two users may decide differently, for example, on whether or not to include a small trail in the map. While the first user may consider the trail unimportant, the second user may recognise the trail as an important link in a network of hiking paths. For the success of a map matching algorithm, the inclusion of a short road segment in the data can be crucial. If the trajectory follows a road that is missing in the data, many map matching algorithms introduce long detours (Fig. 5.4). Similar

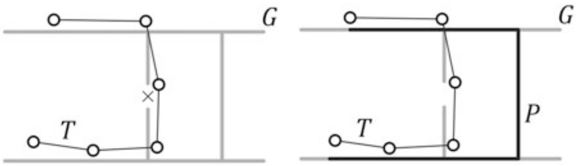

Fig. 5.4 Given a geometric graph $\boldsymbol{G}$ (the road network) and a polyline $\boldsymbol{T}$ (the GPS trajectory), the topological map matching algorithm of Alt et al. (2003), computes a path $\boldsymbol{P}$ in $\boldsymbol{G}$ that has the minimum Fréchet distance to $\boldsymbol{T}$. (Assuming that a dog runs along $\boldsymbol{T}$ and a man runs along $\boldsymbol{P}$ such that neither the dog nor the man are allowed to walk backwards, the Fréchet distance between $\boldsymbol{T}$ and $\boldsymbol{P}$ is the minimum length of a leash that allows the dog and the man to be kept connected.) Because the road network contains a small gap (*left*, marked with $\times$), the output path $\boldsymbol{P}$ (*right*) is much longer than $\boldsymbol{T}$

problems occur in the case of topological errors, which are often caused by users who add new paths without carefully integrating them into the present data.

In this Case study, we discuss experiments with a topological map matching algorithm that has been specifically designed to cope with incomplete road data (Haunert and Budig 2012). We review this algorithm and discuss experiments with user-generated road data from the OpenStreetMap (OSM) project and GPS trajectories recorded during four hikes.

5.7.2 Algorithm

Our map matching algorithm for incomplete road data (Haunert and Budig 2012) extends a basic map matching algorithm by Newson and Krumm (2009). For each point p_i of the GPS trajectory $p_1, \ldots, p_n$, the basic algorithm first selects a discrete set of k candidate matches $C_i = \left\{c_i^1, \ldots, c_i^k\right\}$, where k can be set by the user. Each candidate match is a point in a geometric graph G, which represents the road network. The output path is determined by selecting one candidate c_i^* of each set C_i and, for $i = 1, \ldots, n-1$, connecting c_i^* and c_{i+1}^* via a shortest path in G. Figure 5.5, left, illustrates this approach.

In order to select a sequence of candidates of maximum likelihood, Newson and Krumm (2009) apply a Hidden Markov Model (HMM). This requires that for each candidate match $c \in C_i$ an *observation probability density* is defined, that is, the probability density that GPS point p_i is observed if the actual position of the user is c. Furthermore, for each two candidates $c \in C_i$ and $d \in C_{i+1}$ a *transition probability* is defined, which models the probability that a user in c moves (within the time between two GPS measurements) to d. In the model by Newson and Krumm (2009), the probability of a transition from a candidate c to a candidate d is a function of the graph distance (the length of the shortest path) from c to d in G. Given the sets of candidate matches, the observation probabilities, and the

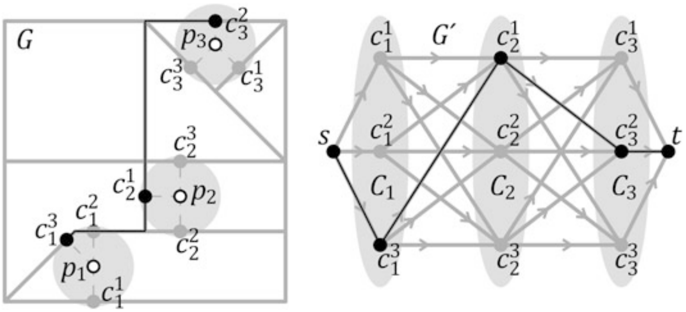

Fig. 5.5 *Left* The graph $\boldsymbol{G}$ representing the road network, three points forming a GPS trajectory $\langle \boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3 \rangle$ three candidate matches for each GPS point, and the output path (*black*) resulting from a selection of candidate matches. *Right* The directed graph $\boldsymbol{G}'$ for the problem instance in the *left* figure. Each $\boldsymbol{s}$-$\boldsymbol{t}$-path in $\boldsymbol{G}'$ corresponds to a solution to the map matching problem. The Viterbi algorithm selects the $\boldsymbol{s}$-$\boldsymbol{t}$-path that maximises the product of arc weights

transition probabilities, a weighted directed graph G' is defined, which is sketched in Fig. 5.5, right (G' contains a node for each candidate plus two dummy nodes s and t). Then, the Viterbi algorithm (Rabiner and Juang 1986) is applied to find the s-t-path in G' that maximises the product of arc weights. Given that the weights of the arcs of G' are appropriately defined (based on the observation probability densities and transition probabilities) this path can be interpreted as the sequence of candidate matches that explains the given GPS observations best.

Our map matching algorithm for incomplete road data (Haunert and Budig 2012) also uses the idea of a discrete set of candidate matches. The difference between our method and the method of Newson and Krumm is, however, that we include one additional candidate in the candidate set C_i of each GPS point p_i. We define this candidate as having the same position as p_i, thus it does not necessarily lie on an edge of the road data set. Therefore, we term such a point an *off-road candidate* and thereby distinguish it from *on-road candidates* that lie on edges of the road network. Our idea is to avoid long detours by allowing a GPS point to be matched with its off-road candidate. If we select the off-road candidate from a candidate set C_i and the off-road candidate from the candidate set C_{i+1}, we define that the output path contains the straight-line segment connecting both candidates. If, on the other hand, we select two on-road candidates for two consecutive GPS points, then our algorithm (just as the basic algorithm by Newson and Krumm) connects both candidates with a shortest path in G. In the case that for any two consecutive GPS points one on-road candidate and one off-road candidate are selected, we introduce a path connecting both candidates by concatenating a path in G with a straight-line segment. Among all possible concatenations, we choose the one of minimum total cost, where we define the cost of the path in G to equal its length and the cost of the straight-line segment to equal the product of its length and a constant factor. To avoid a situation in which the algorithm only selects off-road candidates, we define the transition probability from any node to an on-road

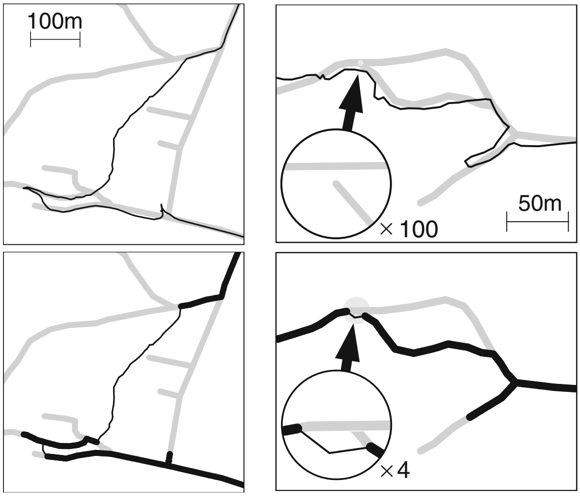

Fig. 5.6 *Left* A trajectory (*top*, *black*) with a road network (*gray*) and the output path (*bottom*, *black*) with on-road parts (*bold*) and off-road parts (*thin*) yielded by our algorithm. *Right* A sample from our data with a topological error (see close-up view in *circle*, *top*). The map matching algorithm selects the off-road candidate of one GPS point and thereby avoids a long detour

candidate to be higher (by a constant factor) than the transition probability to an off-road candidate.

5.7.3 Experimental Results

We implemented our algorithm in Java and tested it with OSM data and trajectories that we recorded with a GPS receiver of type Garmin Edge 710 during four hikes in the surroundings of Würzburg, Germany. In total, the GPS receiver yielded 7,799 GPS points over a distance of 65 km (on average, one GPS point every eight meters). In order to orient ourselves, we used a tourist map of scale 1:75,000 and the GPS receiver with a digital map based on the OSM data. In some cases, we took shortcuts along small trails, but we mainly used official hiking paths. We conducted experiments to find a good set of parameters, which we have used by default since. As a result of our test, we found 23 off-road sections in our output paths, that is, sections where we did not use edges of the road network. Figure 5.6 shows two examples with off-road sections.

To evaluate the robustness of our algorithm, we conducted additional tests in which we removed a randomly selected set of edges from the original road data. In the paths that we obtained by matching the original road data with two of our tracks (one in a rural and one in an urban area), we did not find significant errors and therefore used them as ground truth for our tests.

In each of our tests, we randomly selected and removed a certain percentage of the edges from our original road data set and matched one of the trajectories with the remaining roads. We term the percentage of the edges that we kept in the road network the *integrity* of the road network. Using this approach, we solved more

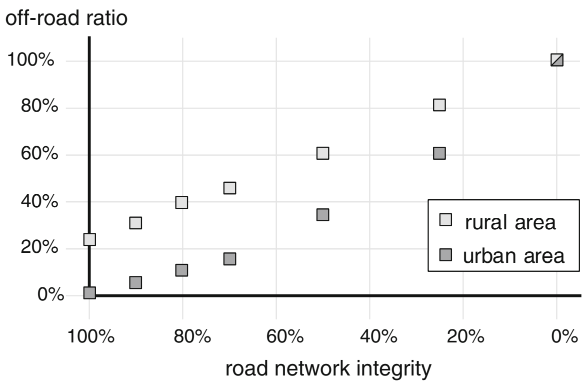

Fig. 5.7 Influence of the road network integrity on the off-road ratio

than 3,000 test instances with networks of different integrity. For each output path, we computed the *off-road ratio*, that is, the ratio between the length of all off-road parts of the path and its total length. Figure 5.7 summarises the results by showing averages over instances with the same trajectory and road networks of the same integrity. We observe that, as the integrity declines, generally more off-road edges are used in the output path, but we also observe differences between the results for the urban area and the rural area. In the urban area, the algorithm often used alternative road edges for edges we had removed from the road data. This is because the trajectory passed a public park with many mutually parallel paths. For example, the data set contained a bike track and a path for pedestrians next to each other. In contrast, for our track in the rural area, the algorithm seldom chose an alternative on-road route for a road we had removed from the road data. Instead, an off-road edge was introduced. In all cases, the total length of the output trajectory increased only marginally, for example, by a mere of 2 % for the urban data set of 80 % integrity.

5.7.4 Conclusions

Topological map matching algorithms often fail if applied to incomplete road data. In our experiments with OSM data and four GPS trajectories, we revealed 23 situations in which it was necessary to include off-road sections to the output paths, otherwise long detours would have been needed. Therefore, the problem cannot be neglected when dealing with user-generated data. With only a small modification of the existing map-matching algorithm by Newson and Krumm (2009), however, all four real-world instances on which we tested our algorithm were correctly solved. Even when we reduced the integrity of the road network drastically, our algorithm produced satisfactory results, but in urban areas sometimes matched the trajectory to a path that was parallel to the correct path. As an effect, when reducing the integrity of the road network to a certain percentage, the ratio between the length of all on-road sections and the total length of the path was reduced to a slightly lower percentage. Unlike most of the existing map matching algorithms, however, our algorithm does not require the output path to be fully contained in the road data and thus avoids long detours. Our experiment in which we removed a random set of road segments from the road data also suggests that the algorithm will cope with generalised data in which minor roads have been removed. In order to better assess the effect of map generalisation on the result of a map matching algorithm, however, further research is needed. For example, it would be interesting to investigate how line simplification affects the quality of the output path.

5.8 Conclusions

Visual presentation of the new (mass) data available gathered from human sensors and sensor networks poses a challenge for cartography. This new data source marks a fundamental change from mapping as an art performed by experts who produce maps as an end product for the general public—to the new situation where people both generate data and want to access it immediately. This implies that visualisation and map generalisation need to be simple to apply and fast. Also, mechanisms have to be available to clean the data due to frequent errors and redundancy. Furthermore, as most VGI is acquired as point data, the integration with known spatial data sets (topographic data) is needed, in order to provide the correct spatial context.

We conclude that the visualisation of VGI has great potential and provides great opportunities, but also presents us with some challenges. In order to exploit the benefits of the data, the following issues have to be tackled:

- (New) design principles to visualise VGI have to be developed, especially to visualise them in conjunction with background maps;
- Methods for automatically detecting and interpreting meta-information of VGI data are required. Meta information relate to semantics, scale, quality, timeliness, coverage, and consistency;
- Methods to visualise this meta-information have to be developed and made available;
- Generalisation methods for VGI have to be developed and made available. Especially methods for selection, aggregation and dimensional change, as they are the most relevant for frequently occurring point data;
- Data integration methods have to be developed, which take into account meta-data, known semantics and relations to background objects.

References

Alt H, Efrat A, Rote G, Wenk C (2003) Matching planar maps. In: Proceedings of the fourteenth annual ACM-SIAM symposium on discrete algorithms (SODA '03). Society for industrial and applied mathematics. Philadelphia, PA, USA, pp 589–598

Atkinson GM, Wald DJ (2007) "Did you feel it?" intensity data: a surprisingly good measure of earthquake ground motion. Seismol Res Lett 78(3):362–368

Balley S, Parent C, Spaccapietra S (2004) Modelling geographic data with multiple representations. Int J Geogr Inf Sci 18(4):327–352

Bereuter P, Weibel R and Burghardt D (2012) Content zooming and exploration for mobile maps. In: Proceedings 15th AGILE international conference on geographic information science, Avignon, France

Bobzien M, Petzold I, Burghardt D (2007) Automated derivation of a 1:300,000 topographic map from swiss dlm vector 200. In: Proceedings of the international cartographic congress, Moscow, Russia, International cartographic association

Brassel KE, Weibel R (1988) A review and conceptional framework of automated map generalization. Int J Geogr Inf Syst 2(3):224–229

Bröring A, Echterhoff J, Jirka S, Simonis I, Everding T, Stasch C, Liang S, Lemmens R (2011) New generation sensor web enablement. Sens Open Access J 11(3):2652–2699

Cao L, Krumm J (2009) From GPS traces to a routable road map. In: 17th ACM SIGSPATIAL International conference on advances in geographic information systems (ACM SIGSPATIAL GIS 2009), November 4-6, 2009. Seattle, WA, pp 3–12

Carver S (2001) Public participation using web-based GIS. Environ Plan B: Plan Des 28(6):803–804. doi:10.1068/b2806ed

Chen Y, Krumm J (2010) Probabilistic modeling of traffic lanes from GPS traces. In: 18th ACM SIGSPATIAL international conference on advances in geographic information systems (ACM SIGSPATIAL GIS 2010), November 2-5, 2010. San Jose, CA

Codescu M, Horsinka G, Kutz O, Mossakowski T, Rau R (2011) OSMonto—an ontology of OpenStreetMap tags. http://www.informatik.uni-bremen.de/~okutz/osmonto.pdf. Accessed 27 Feb 2013

Coleman DJ, Geogiadou Y, Labonte J (2009) Volunteered geographic information: the nature and motivation of producers. Int J Spat Data Infrastruct Res 4:332–358

Dahinden T, Sester M (2009) Categorization of linear objects for map generalization using geocoded articles of a knowledge repository. In: Proceedings of the international workshop presenting spatial information: granularity, relevance, and integration, held in conjunction with the conference on spatial information theory, COSIT

Davies JJ, Beresford AR, Hopper A (2006) Scalable, distributed, real-time map generation. IEEE Pervasive Comput 5(4):47–54

Dalyot S, Dahinden T, Schulze, MJ, Boljen, J, Sester, M (2012) Geometrical adjustment towards the alignment of vector databases. In: ISPRS annals of photogrammetry, remote sensing and spatial information sciences, vol. I–4, pp. 13–18

Devillers R, Jeansoulin R (2006) Fundamentals of spatial data quality. ISTE, London

Devillers R, Bédard Y, Jeansoulin R, Moulin B (2007) Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. IJGIS 21(3):261–282

Devogele T, Trevisan J, Raynal L (1996) Building a multiscale database with scale-transition relationships. In: Proceedings of the 7th international symposium on spatial data handling, advances in GIS research II. Delft, The Netherlands, pp 6.19–6.33

Duckham M (2012) Decentralized spatial computing: foundations of geosensor networks. Springer, Berlin

Fink M, Haunert JH, Schulz A, Spoerhase J, Wolff A (2012) Algorithms for labeling focus regions. Vis Comput Graph IEEE Trans 18(12):2583–2592

Fritz S, McCallum I, Schill C, Perger C, See L, Schepaschenko D, Van der Velde M, Kraxner F, Obersteiner M (2012) Geo-Wiki: an online platform for improving global land cover. Environ Model Softw 31:110–123

Gartner G (2012) openemotionmap.org—emotional response to space as an additional concept in cartography, international archives of the photogrammetry, remote sensing and spatial information sciences, vol XXXIX-B4

Gervais M, Bedard Y, Levesque M, Bernier E, Devillers R (2009) Data quality issues and geographic knowledge discovery, pp 99–116

Girres JF, Touya G (2010) Quality assessment of the French OpenStreetMap dataset. Trans GIS 14(4):435–459

Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. GeoJournal 69:211–221

Goodchild MF, Li L (2012) Assuring the quality of volunteered geographic information. Spat Stat 1:110–120

Grünreich D (1992) ATKIS—a topographic information system as a basis for GIS and digital cartography in Germany. In: Vinken R (ed) From digital map series in geosciences to geo-information systems. Geologisches Jahrbuch, Federal institute of geosciences and resources, Hannover, vol A(122), pp 207–216

Guptill SC, Morrison J (1995) Elements of spatial data quality. Elsevier Science, New York

Hahmann S, Burghardt D (2010) Linked data—a multiple representation database at web scale? In: Proceedings of the 13th ICA workshop on generalisation and multiple representation, Zurich, Switzerland

Haklay M (2010) How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. Environ Plann B: Plann Des 37(4):682–703

Haklay M, Basiouka S, Antoniou V, Ather A (2010) How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. Cartographic J 47(4):315–322

Hampe M, Sester M, Harrie L (2004) Multiple representation databases to support visualisation on mobile devices, international archives of photogrammetry and remote sensing, vol 35. Istanbul, Turkey

Haunert JH, Budig B (2012) An algorithm for map matching given incomplete road data. In: Proceedings of the 20th ACM SIGSPATIAL international conference on advances in geographic information systems (ACM-GIS'12), pp 510–513

Heipke C (2010) Crowdsourcing geospatial data. ISPRS J Photogrammetry Remote Sens 65(6):550–557. doi:10.1016/j.isprsjprs.2010.06.005

Helbich M, Amelunxen C, Neis P (2012) Comparative spatial analysis of positional accuracy of openstreetmap and proprietary geodata. Int. GI_Forum 2012. Salzburg, Austria

Henson C, Pschorr J, Sheth A, Thirunarayan K (2009) SemSOS semantic sensor observation service. International symposium on collaborative technologies and systems

Hollenstein L, Purves R (2010) Exploring place through user-generated content: using Flickr tags to describe city cores. J Spat Inf Sci 1(2010):21–48

Jaara K, Duchêne C, Ruas A (2011) Toward the generalisation of cartographic mashups: taking into account the dependency between the thematic data and the reference data throughout the process of automatic generalisation, 14th ICA workshop on generalisation and multiple representation, Paris, France, 30 June–1 July 2011

Jokar Arsanjani J, Barron C, Bakillah M, Helbich M (2013a) Assessing the quality of OpenStreetMap contributors together with their contributions. In: Proceedings of the 16th AGILE conference, Leuven, Belgium

Jokar Arsanjani J, Helbich M, Bakillah M, Hagenauer J, Zipf A (2013b) Towards mapping land use patterns from volunteered geographic information. Int J Geogr Inf Sci. doi:10.1080/13658816.2013.800871

Jones CB, Purves P, Clough D, Joho H (2008) Modelling vague places with knowledge from the web. Int J Geogr Inf Sci 22(10):1045–1065

Kieler B, Huang W, Haunert JH, Jiang J (2009) Matching river datasets of different scales. Advances in GIScience. In: Proceedings of 12th AGILE conference on GIScience, lecture notes in Geoinformation and cartography, Springer, Berlin, pp 135–154

Kilpeläinen T (1992) Multiple representations and knowledge-based generalization of topographical data. In: Waugh TC, Healey RG (eds) Advances in GIS research. In: Proceedings of the 6th international symposium on spatial data handling, IGU commission on GIS and association for geographic information, Edinburgh, GB, pp 882–893

Kinkeldey C, Schiewe J (2012) Visualisierung thematischer Unsicherheiten mit "noise annotation lines". Kartographische Nachrichten 62(5):241–249

Lüscher P, Weibel R (2013) Exploiting empirical knowledge for automatic delineation of city centres from large-scale topographic databases. Comput Environ Urban Syst 37:18–34. doi:10.1016/j.compenvurbsys.2012.07.001

MacEachren AM (1992) Visualizing uncertain information. In: cartographic perspectives, vol 13, p 10

Mackaness W, Ruas A, Sarjakoski T (2007) Generalisation of geographic information: cartographic modelling and applications. Oxford, Elsevier

Mackaness WA, Chaudhry O (2013) Assessing the veracity of TF-IDF for extracting place semantics from Flickr tags transactions in GIS (in press)

McMaster RB, Shea KS (1992) Generalization in digital cartography. Association of American Geographers, Washington

Mondzech J, Sester M (2011) Quality analysis of OpenStreetMap data based on application needs. Cartographica 46(2):115–125

Mustière S, Devogele T (2008) Matching networks with different levels of detail. GeoInformatica 12(4):435–453

Neis P, Zielstra D, Zipf A (2012) The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. Future Internet 4(1):1–21

Newson P and Krumm J (2009) Hidden markov map matching through noise and sparseness. In: Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems (GIS '09), pp 336–343

Noy N (2004) Semantic integration: a survey of ontology based approaches. SIGMOD Rec 33:65–70

Paelke V, Dahinden T, Eggert D, Mondzech J (2012) Location based context awareness through tag-cloud visualizations. Adv Geo-Spatial Inf Sci (CRC Press)

Poser K, Dransch D (2010) Volunteered geographic information for disaster management with application to rapid flood damage estimation. Geomatica 64(1):89–98

Quddus MA, Ochieng WY, Noland RB (2007) Current map-matching algorithms for transport applications: state-of-the art and future research directions. Transp Res Part C: Emerg Technol 15(5):312–328

Rabiner LR, Juang BH (1986) An introduction to hidden Markov models. IEEE ASSP Mag 3(1):4–15

Rak A, Coleman D, Nichols S (2012) Legal liability concerns surrounding volunteered geographic information applicable to Canada. In: Rajabifard, A, Coleman D (eds) Spatially enabling government, industry and citizens, USA: GSDI Association Press, pp 125–143. http://www.gsdi.org/gsdiconf/gsdi13/papers/256.pdf

Rambaldi G, Mccall M, Weiner D, Mbile P, Kyem P (2004) Participatory GIS. http://www.iapad.org/participatory_gis.htm

Ramm F (2012) Am Puls der Zeit—Minütliche updates bei OpenStreetMap. Paper presented at the FOSSGIS 2012, Dessau

Ramm F, Topf J, Chilton S (eds) (2010) OpenStreetMap: using and enhancing the free map of the world. UIT Cambridge, Cambridge

Roick O, Heuser S (2013) Location based social networks—definition, current state of the art and research agenda. Trans GIS. doi:10.1111/tgis.12032

Rouse LJ, Bergeron SJ, Harris TM (2007) Participating in the geospatial web: collaborative mapping, social networks and participatory GIS. In: Scharl A, Tochtermann K (eds) The

Geospatial Web. Springer, London, pp 153–158. Retrieved from http://dx.doi.org/10.1007/978-1-84628-827-2_14
Safra E, Doytsher Y (2006) Integration of multiple geo-spatial datasets. ASPRS 2006 annual conference, Reno, Nevada
Samet H (eds) (2005) Foundations of multidimensional and metric data structures. The Morgan Kaufmann series in computer graphics and geometric modeling. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA
Sayda F (2005) Involving LBS users in data acquisition and update. In: Proceedings of the AGILE, Portugal
Schroedl S, Wagstaff K, Rogers S, Langley P, Wilson C (2004) Mining GPS traces for map refinement. Data Min Knowl Disc 9(1):59–87
Scofield RP, Christie D, Sagar PM, Sullivan B (2012) eBird and avifaunal monitoring by the ornithological society of New Zealand. NZ J Ecol 36:279–286
Senaratne H, Gerharz L (2011) An assessment and categorisation of quantitative uncertainty visualisation methods. The 14th AGILE international conference on geographic information science, Utrecht, Netherlands, 18–21 April 2011
Sengstock C, Gertz M (2012) Latent geographic feature extraction from social media. In: Proceedings of the 20th ACM SIGSPATIAL international conference on advances in geographic information systems, Redondo Beach
Sheeren D, Mustière S, Zucker JD (2009) A data-mining approach for assessing consistency between multiple representations in spatial databases. Int J Geogr Inf Sci 23(8):961–992
Siebritz L, Sithole G, Zlatanova S (2012) Assessment of the homogeneity of volunteered geographic information in South Africa. In: international archives of the photogrammetry, remote sensing and spatial information sciences, vol 39-B4. XXII ISPRS Congress, Melbourne, Australia
Siriba DN, Dalyot S, Sester M (2012) Geometric quality enhancement of legacy graphical cadastral datasets through thin plate splines transformation. Surv Revi 44(325):91–101
Slingsby A, Dykes J, Wood J, Clarke K (2007) Mashup cartography: cartographic issues of using Google earth for tag maps. ICA commission on maps and the internet, ICA, Warsaw, Poland, pp 79–93
Smart PD, Quinn JA, Jones CB (2011) City model enrichment. ISPRS J Photogram Remote Sens 6(2):223–234
Stefanidis A, Nittel S (2004) Geosensor networks. CRC Press, Boca Raton
Stigmar H (2005) Matching route data and topographic data in a real-time environment. In: Hauska H, Tveite H (eds) ScanGIS'2005—proceedings of the 10th scandinavian research conference on geographical information sciences, Stockholm, Sweden, pp 89–107
Stoter JE, Meijers M, van Oosterom P, Grünreich D, Kraak MJ (2010) Applying DLM and DCM concepts in a multi—scale data environment. Paper presented at GDI 2010: a symposium on generalization and data integration, Boulder, USA, 20–22 June 2010
Sullivan BL, Wood CL, Iliff MJ, Bonney RE, Fink D, Kelling S (2009) eBird: a citizen-based bird observation network in the biological sciences. Biol Conserv 142(10):2282–2292
Surowiecki J (ed) (2004) The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies. Societies and Nations Little, Doubleday
Tauscher S, Neumann K (2012) Combining web map services and opinion mining to generate sentiment maps of touristic locations, symposium on service oriented mapping, Wien, JOBSTMedia Präsentation Management Verlag, pp 277–286, 11/2012
Touya G (2012) What is the level of detail of OpenStreetMap? Role of volunteered geographic information in advancing science: quality and credibility. In: conjunction with GIScience 2012, 18 Sept 2012
van Oort P (2006) Spatial data quality: from description to application. PhD Thesis, Wageningen University, The Netherlands
Volz S (2005) Data-driven Matching of Geospatial Schemas. In: Cohn AG, Mark DM (eds) Spatial information theory. Proceedings of the international conference on spatial information theory (COSIT '05), Ellicottville, NY. Lecture Notes in Computer Science 3693, pp. 115–132

Walter V, Fritsch D (1999) Matching spatial data sets: a statistical approach. Int J Geogr Inf Sci 13(5): 445–473

Warneke H, Schäfers M, Lipeck U, Bobrich J (2011) Matching-based map generalization by transferring geometric representations. In: Proceedings Geoinformatik 2011, Münster, pp 71–77

Weibel R, Dutton G (2005) Generalizing spatial data and dealing with multiple representations. In: Longley P, Goodchild MF, Maguire DJ, Rhind DW (eds) Geographical information systems: principles, techniques, management and applications, 2nd edn (Abridged Edition). Wiley

Yates D, Paquette S (2011) Emergency knowledge management and social media technologies: a case study of the 2010 Haitian earthquake. Int J Inf Manage 31(1):6–13

Yuan S, Tao C (1999) Development of conflation components. In: Proceedings of the international conference on geoinformatics and socioinformatics. Ann Arbor, MI, pp 1–12. http://www.umich.edu/yiinet/chinadata/geoim99/Proceedings/yuan_shuxin.pdf

Zhang Y, Yang B, Luan X (2012) Automated matching crowdsourcing road networks using probabilistic relaxation. In: ISPRS annals of photogrammetry, remote sensing and spatial information sciences, vol 1–4

Zook M, Graham M, Shelton T, Gorman S (2010) Volunteered geographic information and crowdsourcing disaster relief: a case study of the haitian earthquake. World Med Health Policy 2:7–33. doi:10.2202/1948-4682.1069