Transactions in GIS

Research Article

Transactions in GIS, 2014, 18(6): 877–895

A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis

Christopher Barron, Pascal Neis and Alexander Zipf

Department of Geography, University of Heidelberg

Abstract

OpenStreetMap (OSM) is one of the most popular examples of a Volunteered Geographic Information (VGI) project. In the past years it has become a serious alternative source for geodata. Since the quality of OSM data can vary strongly, different aspects have been investigated in several scientific studies. In most cases the data is compared with commercial or administrative datasets which, however, are not always accessible due to the lack of availability, contradictory licensing restrictions or high procurement costs. In this investigation a framework containing more than 25 methods and indicators is presented, allowing OSM quality assessments based solely on the data's history. Without the usage of a reference data set, approximate statements on OSM data quality are possible. For this purpose existing methods are taken up, developed further, and integrated into an extensible open source framework. This enables arbitrarily repeatable intrinsic OSM quality analyses for any part of the world.

1 Introduction

In the past decade a significant transition within the World Wide Web (WWW) was carried out leading to an altered usage of the WWW where users no longer act as sheer consumers of predefined content. Instead, they are more and more part of a contributing process, sharing knowledge and information (O'Reilly 2005). Popular examples are the Internet encyclopedia Wikipedia and content sharing platforms such as Flickr for photos and Youtube for videos. These platforms provide the opportunity of contributing various types of content, so-called User-Generated Content (UGC) (Brando and Bucher 2010; Chilton 2009; Goodchild 2009). Besides, Volunteered Geographic Information (VGI) can be thought of as a special case of UGC. VGI, also referred to as crowd-sourced geodata, is defined as the collaborative acquisition of geographical information and local knowledge by volunteers, amateurs or professionals (Goodchild 2007). Among others, e.g. Map Insight (http://mapinsight.teleatlas.com), Map Reporter (http://mapreporter.navteq.com/), Wikimapia (http://wikimapia.org/) or Google Map Maker (http://www.google.com/mapmaker), OpenStreetMap (OSM) has evolved to one of the greatest and most famous VGI projects in the past years (Chilton 2009; Goodchild and Li 2012) with 1.3 million users registered at August 2013 (OpenStreetMap 2013e). Since commonly no authorized instance examines the contributed information, data quality assurance plays a crucial role within the OSM project (Flanagin and Metzger 2008; Goodchild and Li 2012). This fact is becoming more and more important, not least because OSM turns out to be a serious geodata alternative for different applications and is used in a wide range of geographic information systems (GIS) and applications (Amelunxen 2010; Goetz and Zipf 2013; Hagenauer and Helbich 2012).

Address for correspondence: Christopher Barron, Department of Geography, University of Heidelberg, Berliner Strasse 48, D-69120 Heidelberg, Germany. E-mail: christopher.barron@geog.uni-heidelberg.de

doi: 10.1111/tgis.12073

A commonly used way of assessing the OSM data quality is the comparison with ground truth reference datasets (Girres and Touya 2010; Haklay 2010; Helbich et al. 2012; Mooney et al. 2010a; Neis et al. 2012; Zielstra and Hochmair 2012). However, accessibility to high quality and commercial datasets for such extrinsic analyses is often limited due to costs and licensing restrictions (Mooney et al. 2010a). Therefore, suitable alternatives are necessary. The key motivation for this article is to investigate how OSM data can be evaluated without a reference for comparison purposes. One possible approach is the investigation of the data's history (van Exel et al. 2010). From the data's history, so-called intrinsic indicators present one opportunity to supply information regarding the data quality. For this purpose, however, new methods, indicators and visualization types are needed to evaluate the quality of OSM data. To this end, a framework for intrinsic OSM data quality analyses, named *iOSMAnalyzer*, was developed. The framework was implemented as a tool using free and open source components. This allows anyone to generate information about OSM data quality for a freely selectable area using only OSM's data history. In the context of spatial data quality analysis, Devillers et al. (2002) discussed the limitations of metadata as an assessment factor to help users to evaluate if a dataset is usable or not. These findings resulted in the introduction of a system which proved that it is relevant to include data quality visualization issues in the communication between producers and users of the data (Devillers et al. 2002, 2007). Thus an additional motivation was that the developed framework should facilitate the decision whether the quality of OSM data in a selected area of a user's choice is sufficient for her or his use case or not.

The remainder of this article is organized as follows. Section 2 gives an introduction to OSM and summarizes related scientific studies on OSM data quality. The main focus in this work is on Section 3. Before a short introduction to the developed framework, several methods and indicators for intrinsic OSM quality analyses are introduced. Section 4 evaluates the outcome using exemplary results of different regions. Finally, Section 5 summarizes the results of this investigation and discusses further research needs.

2 The OpenStreetMap Project: Introduction and Related State of the Art Research

The goal of the OSM project is to create a free and editable world map (Ramm et al. 2011). Within the project volunteers, amateurs and professionals from different social worlds (Lin 2011) act as sensors (Flanagin and Metzger 2008) and collect geographic data. This bottom up process stands in contrast with the traditional centralized procedure of collecting geographic data (Goodchild 2007). The motivation for contributing to OSM varies heavily: it ranges from self-expression over manifestation and representation of people's online identity to a simple fun factor. Meaningful extracurricular activities, interesting technologies and a fascinating general project development are further motivational reasons (Budhathoki 2010). In general, data for OSM can be derived from multiple sources and edited and imported by means of different freely available editors. The most popular editors are the Java OpenStreetMap Editor (JOSM) (http://josm.openstreetmap.de/), the online flash editor Potlatch (http://www.openstreetmap.org/edit?editor=potlatch2) or the web-based JavaScript editor iD (http://ideditor.com/). The classic approach is the collection of spatial data with portable and GPS enabled devices. In addition, several companies such as Aerowest, Microsoft Bing (Bing 2010) or Yahoo! released, at least temporarily, their aerial images for the OSM project. The community is allowed to use these images as a base layer for tracing geographic features, such as for example buildings, forests or lakes. The contributors' local knowledge is also a valuable source of geographic

information. Furthermore, datasets, which fit the licensing restrictions, can also be imported to the OSM database. At best, this is done in close collaboration with the (local) community and respective mailing lists as the appropriateness of imports is discussed controversially (Zielstra et al. 2013).

All contributed data is stored according to the OSM data model wherein point features are represented by "Nodes" and linear features by "Ways". Polygonal objects are represented by "closed Ways". Additionally, features can be further specified semantically by key-value pairs, so-called tags. There are no restrictions to the usage of tags. Whereas traditional authoritative and commercial data sets usually follow the Resource Description Framework (RDF) notion, each OSM feature can hold multiple tags or no tags at all. Nevertheless, for the purpose of consistency, it is recommended to use commonly accepted key-value pairs from the OSM map features web page (OpenStreetMap 2013b). Finally, Relations are used to model logical relationships between the previously mentioned features (Ramm et al. 2011).

2.1 Parameters of Geodata Quality

Quality in general plays a key role when working with all kinds of geodata, especially in data production and assessment (Veregin 1999) or exchange (Goodchild 1995). This is especially the case with OSM data, as the contributors are not faced with any restrictions during the data collection and annotation process. In the field of geo-information, the principles of the "International Organization for Standardization" (ISO) can be taken into account for quality assessment. The ISO 19113 standard describes general principles of geodata quality (http://www.iso.org/iso/catalogue_detail.htm?csnumber=26018) and ISO 19114 contains procedures for quality evaluation of digital geographic datasets (http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=26019). The ISO 19157 "Geographic Information: Data Quality" standard (http://www.iso.org/iso/catalogue_detail.htm?csnumber=32575), currently under development, aims to harmonize all standards related to data quality and revises the aforementioned standards. The quality of spatial data can be evaluated with the help of following elements of ISO 19113:

- "Completeness": describes how complete a dataset is. A surplus of data is referred to as "Error of Commission", a lack of data in contrast as "Error of Omission".
- "Logical Consistency": declares the accuracy of the relations manifested within a dataset. This element can be further subdivided into "intra-theme consistency" and "inter-theme consistency".
- "Positional Accuracy": defines the relative and absolute accuracy of coordinate values.
- "Temporal Accuracy": the historical evolution of the dataset.
- "Thematic Accuracy": describes the accuracy of the attributes assigned to a geometry.

However, OSM data quality heavily depends on the purpose for which the data will be deployed. We refer to this as "Fitness for Purpose" assessment, previously defined by Veregin (1999) as determining "fitness-for-use".

2.2 Quality Assessment in OpenStreetMap – Overview of Related Scientific Research

The increasing availability of voluntarily and collaboratively collected geodata, in particular OSM, led to numerous scientific studies with a focus on the evaluation of this data. In the beginning, investigations mainly focused on the OSM road network with the help of a ground truth reference dataset. For instance, Haklay (2010) compared the OSM road network with

Ordnance Survey Meridian 2 for England and Kounadi (2011) with the Hellenic Military Geographical Service (HMGS) dataset for Athens, Greece. For Germany, Zielstra and Zipf (2010) conducted a comparison with TeleAtlas-MultiNet, Ludwig et al. (2011) with Navteq and, for the period from 2007 until 2011, Neis et al. (2012) with TomTom MultiNet. Using a different method, Helbich et al. (2012) also investigated the positional accuracy of the OSM road network. Employing a spatial statistical comparison method, the authors compare identical road junctions with TomTom and official survey data as a reference. All previously mentioned studies on the OSM road network show, broadly speaking, one commonality: a high positional accuracy and a huge amount of details are found around urban areas with a high number of contributors. In contrast, more rural areas often show a lower level of OSM data quality. However, some urban areas, Istanbul for example, show a high number of contributions by mappers with their main activity area located more than 1,000 km away (Neis et al. 2013). The authors state that in some way this mapping behavior contradicts the original idea of VGI as projects where people contribute their local knowledge.

Beside the road network, other features of OSM have also been object of interest for quality investigations. Mooney et al. (2010b) compare OSM land cover features with the Ordnance Survey Ireland (OSI) dataset using shape similarity tests. Girres and Touya (2010) examine different quality aspects of the French OSM dataset according to quality principles stated in ISO 19113: 2002. The authors highlight the problem of heterogeneity in VGI datasets such as OSM which is, among others, caused by the contributors' freedom within the data collection process.

As mentioned, a considerable number of studies have evaluated different data quality aspects of OSM data. In most cases, data is evaluated and compared with authoritative datasets. On the other hand, very few studies target analyses conducted without a reference dataset. However, as stated by Batini and Scannapieco (2006), intrinsic data quality analyses capture the data's inherent quality and according to Wang and Strong (1996) and Batini and Scannapieco (2006) it includes the following dimensions: accuracy, objectivity, believability and reputation. In the case of OSM, Mooney and Corcoran (2012a) attempted to assess the quality of OSM features by analyzing objects with more than 15 different versions ("heavily edited objects") for several countries utilizing an OSM-Full-History-Dump. In another investigation the authors examine the tag assignment and the influence of the number of contributors on it (Mooney and Corcoran 2012b). However, the results of both studies showed that the number of contributors to an object does not necessarily relate to the number of tags of an OSM feature. Furthermore, recent studies increasingly dwell on the contributors behind the submitted data. In this context the terminologies "feature quality", "user quality" and their "interdependency" are introduced (van Exel et al. 2010). MVP-OSM, a tool for identifying areas of high quality contributions, also uses this approach. The tool's results are based on the contributors' local knowledge, identifying their experience and community recognition for a selected area (Napolitano and Mooney 2012). Moreover, a conceptual model to analyze contributor patterns in VGI projects in a more structured way is proposed (Rehrl et al. 2013). Following a data-orientated approach, a provenance vocabulary is presented by Keßler et al. (2011) allowing statements on the lineage of OSM data based on the (editing) history. Touya and Brando-Escabor (2013) proposed a method to infer the level of detail of OSM features based on several criteria such as the geometric resolution and the feature type. A model to estimate the uncertainty of geometric measurements of vector objects has been introduced by Girres (2011). For the investigation and estimation of length errors for different kinds of road samples the TOP100 road network of France have been utilized.

However, the overall activity of the contributors within the OSM project is analyzed where the authors illustrate a strong bias in the participation process (Neis and Zipf 2012). It should be noted that only 38% of all registered members have ever accomplished at least on edit and only 5% of all members have contributed in a significant manner. This behavior is closely linked to what is commonly known as the "Participation Inequality" in online communities (Nielsen 2006). The Participation Inequality follows a 90-9-1 pattern and is observed in several communities in the WWW, which rest on contributions of their members. Within those, 90% of all users solely consume information without contributing, 9% contribute from time to time and only 1% is actually responsible for the majority of the content.

2.3 OpenStreetMap Data, History and Tools

OSM data can be obtained from different sources in several file formats. A weekly updated OSM database dump-file (http://planet.openstreetmap.org/) containing a temporal snapshot of the entire world with a current compressed size of approximately 28 GB is available. Smaller extracts of specific regions are also offered by companies (http://download.geofabrik.de/). These files are mainly provided as Esri-shapefiles (*.shp), in a compressed XML (*.osm.bz2) or a Protocolbuffer Binary Format (*.osm.pbf) which makes faster processing possible. Several tools are capable of processing OSM data. Probably the most prominent ones within the OSM software environment are the open source command-line Java tool "osmosis" (http://wiki.openstreetmap.org/wiki/Osmosis) or the flexible C++/JavaScript framework "osmium" (http://wiki.openstreetmap.org/wiki/Osmosis).

Beside the abovementioned snapshot of the recent database, the OSM-Full-History-Dump contains, with minor exceptions, the entire history of the OSM data (http://planet.openstreetmap.org/planet/full-history/). A new version of an OSM object is created whenever a feature's geometry is changed. The simple movement of an already existing Way's Node does not lead to a new version number. Moreover, adding, modifying, or deleting a tag also leads to the increase of a feature's version number. Regarding the versioning, a bug in the Potlatch 1 OSM editor up to 2011 has to be noted. This bug led to an erroneous increase of a feature's version number, although it was not edited, but lay within the spatial extent of an edited OSM changeset (Neis and Zipf 2012). The fact that not every change automatically leads to a new version of an OSM feature and vice versa has to be considered carefully.

Since the introduction of the OSM API 0.5 in October 2007, the recent OSM-Full-History-Dump includes every undertaken addition, modification and deletion within OSM. From contributions during or before OSM API 0.4, only a snapshot of the data which was actually visible at the changeover together with the history of their future changes are available. Moreover, as segments were removed with the introduction of the OSM API 0.5 they are also not included within the OSM-Full-History-Dump. Furthermore, it has to be noted that added or modified data of contributors who did not accept the "Open Database License 1.0" (ODbL) terms during the license change period are also not part of the current OSM-Full-History-Dump any more (OpenStreetMap 2013c).

3 Introducing the Framework

In the following subsections, the developed iOSMAnalyzer framework will be delineated. After an introduction to the applied techniques, several methods and indicators for evaluating OSM

data are presented. Finally, the structure of the developed framework is illustrated. The main focus is on the quality assessment for different "Location Based Services" (LBS) applications.

The approach proposed in this article differes significantly from previously conducted studies in several respects. An OSM-Full-History-Dump is used as a sole input for the analyses. Therefore, with the exception of the aforementioned particularities, the entire temporal dimension of the dataset can be taken into account. Furthermore, no ground truth reference dataset is deployed for OSM data quality evaluation. Therefore, specified areas within OSM can be evaluated regardless of whether a reference is available or not. Thus, a so-called intrinsic analysis approach is applied. For this intrinsic approach, new methods and indicators have been developed because traditional ones from extrinsic analyses are usually only suitable for comparison purposes. Excerpts of numerous techniques presented in this article, for example, are the investigation of the data's historical development, the comparison of features' characteristics at different timestamps or various spatial analyses. In some cases (e.g. feature completeness), however, an intrinsic approach does not allow absolute statements on data quality. Therefore, some results presented with this approach can only act as relative indicators making approximate statements about the possible data quality.

3.1 Defining a Framework for Intrinsic OSM Quality Assessment

As OSM data is used in a wide range of applications, the analyses have to be adjusted to different use cases and specific needs. Hence, in order to evaluate the OSM data, the finally calculated results of the iOSMAnalyzer are divided into the following categories which were selected according to the "Fitness for Purpose" approach: "General Information on the Study Area", "Routing and Navigation", "Geocoding", "Points of Interest-Search", "Map-Applications" and "User Information and Behavior". The fitness of VGI data always depends on the case and needs to be analyzed individually (Mondzech and Sester 2011; Neis et al. 2013). Therefore the framework's results can support the decision whether OSM data could be suitable for one of a great number of use cases. Overall, a set of more than 25 different intrinsic quality indicators (see Figure 1) is considered in the framework. Due to space restrictions a selected number of parameters are presented in the following.

3.2 General Information on the Study Area

The **evolution of OSM features** over a specific period of time provides a first insight into the development and quality of an arbitrarily chosen area within OSM (Neis et al. 2013). For example the cumulated number of contributed points, lines and polygons per month gives a first general and more diverse impression of an area. Histograms allow the visualization of these quantitative developments which have to be interpreted in very different ways: Ciepłuch et al. (2011) allege that OSM datasets rise from the road network. Therefore, first peaks within line evolution histograms could indicate the beginning of general mapping activity. A significantly steeper growth within a few days or weeks is to be expected in case of bulk data imports or automated edits (bots). Both bulk imports and bots are usually documented in the wiki (OpenStreetMap 2013a). Rating these automated edits in terms of good or bad quality heavily depends on the individual case (Zielstra et al. 2013). People also collaboratively contribute to OSM in organized community events. These so-called "mapping parties" possibly lead to significant data increase in a region mainly within a few days potentially enriching existing data (Hristova et al. 2013). The release of aerial images also has an impact on the quantity of data (Neis et al. 2012). For instance, significant peaks of contributed data after

Figure 1 Overview of the iOSMAnalyzer's intrinsic Quality Indicators

December 2010 and in spring 2011 are probably caused by the release of the Bing aerial images for the purpose of digitizing. However, to ensure high attribute quality roads as an example requires local knowledge. Only then their category, name or possible speed limitation can be identified correctly (de Leeuw et al. 2011).

The quantitative calculations mentioned above allow only limited statements on data quality in some cases. As VGI projects such as OSM are mainly driven by their contributors, not only the data but also the behavior of the crowd can be analyzed to provide information about their contributions (Coleman et al. 2009; van Exel et al. 2010; Neis and Zipf 2012; Rehrl et al. 2013). One general indicator is the overall **number of (active) contributors** within an area. Several investigations demonstrate that a high number of active contributors leads to a stable and good quality OSM dataset which is more probably kept up-to-date (Girres and Touya 2010; Haklay et al. 2010; Neis and Zipf 2012). As a consequence, a high and

increasing number of people who have ever created or edited OSM data within an area indicates a possibly better data quality. In addition, the number of actually active contributors per month indicates whether those have contributed only once or in a more frequent way. The higher the number of monthly recurring and contributing mappers, the higher is the heterogeneity of mappers and consequently, the better is the overall data quality. A combination of the general evolution of points, lines or polygons together with the aforementioned information on contributor activity simplifies the interpretation of quantitative feature statistics. Imports and bots are usually carried out by a single registered member leading to a huge amount of created or edited data. By contrast, mapping parties, digitizing from aerial images and simple mapping activities by individuals usually are performed by a high number of contributors.

Beside the overall number of contributors, the **mappers' actual amount of created data** can provide more in-depth information. In a global investigation four different member groups based on the number of created Nodes have been defined by Neis and Zipf (2012): "Senior Mappers" (contributors with 1,000 and more created Nodes), "Junior Mappers" (contributors with at least 10 and less than 1,000 created Nodes), "Nonrecurring Mappers" (contributors with less than 10 created Nodes) and members with no edits. The more mappers with a high number of contributed Nodes can be identified, the more active contributors are present in an area. However, these measurements do not have to be true for a mapper's activity in general. A person identified as a "Nonrecurring Mapper" in one area could be a very active mapper with high contribution rates in another area. Furthermore, there is no evidence that users with a high number of created Nodes also contribute high quality data. This could be expected due to their contribution experience in the selected area; however, this potential thesis requires further research. Concerning the distribution of created or edited features among the mappers, OSM shows an inequality in contributions (Neis and Zipf 2012). As stated by Nielsen (2006) this is referred to as the participation inequality in online communities. The less contributors that are responsible for the major proportion of the data the higher the dependence on those few. These contributors are therefore of particular importance for the OSM project. Moreover, a more uniform distribution shows that more people are contributing and this potentially leads, relatively speaking, to a better overall data quality because errors are more likely detected and fixed.

An important point in OSM is the **currentness of data** (van Exel et al. 2010, Neis and Zipf 2012). After the initial collection process the further maintenance of OSM data is essential for a high quality and up-to-date dataset. Ideally the process of updating the OSM features' geometries and attributes is carried out continuously, homogeneously, throughout and is not limited to specific features. However, this is not the usual case within OSM. A possible way to analyze and represent the currentness is the visualization of the data's latest modification. It can be argued that the last editor of an OSM feature is responsible for its correctness and indirectly confirms this by uploading the modified data to the server. The case is problematic when a feature was already completely and accurately mapped in the past. These features can potentially be detected with the help of adjacent features (van Exel et al. 2010) using a probabilistic approach. Features with an older timestamp surrounded by current features could represent an implicit peer review and attest to their currentness.

The positional accuracy of the OSM data depends very much on the way the data was collected. Several factors such as GPS signal preciseness, displaced aerial images or bulk movements have an impact on data quality. A way to identify these possible positional inaccuracies without a reference dataset is the enhancement and modification of the method proposed by Helbich et al. (2012). Instead of comparing OSM with a ground truth reference dataset, the **location of currently valid road junctions is compared with its previous location**. As already

mentioned, the latter must not necessarily be the last version of the Node representing the junction. By analyzing distance and the angle of two corresponding road junctions within a polar scatter plot, different conclusions regarding the positional accuracy of OSM data can be drawn: on the one hand, an accumulation of points within one angle segment of the diagram indicates possible corrections of the road network caused by a potentially displaced editing basis (either aerial images or GPS traces). Yet a rectification could also be possible but is not clearly distinguishable from deterioration. However, if multiple road junctions show exactly the same distance and angle to their previous location, a bulk movement is very likely. On the other hand, a uniform distribution where all road junctions show an individual distance suggests no positional inaccuracies caused by the abovementioned issues. Within this proposed method, road junctions serve as an indicator. This means that beside the road network other features within the selected dataset could also be affected by positional inaccuracy. Referring to Touya and Brando-Escobar (2013), future research could investigate "*source*" tags or changeset comments of the corresponding features which could provide information about the method of acquisition or the reason for the displacement. The following list contains a summary of the relevant parameters of this section:

- Characterization of active mappers
- Currentness of data
- Evolution of OSM features
- Number of (active) contributors
- Positional accuracy of the OSM (road network) data

3.3 Geodata Quality Assessment for Location Based Services

3.3.1 Routing and navigation

The **completeness of the OSM road network** plays a significant role in routing and navigation applications and has therefore been the subject of several investigations (Girres and Touya 2010; Haklay et al. 2010; Kounady 2011; Ludwig et al. 2011; Neis et al. 2012). In contrast to these thorough comparative studies, this investigation follows an intrinsic approach without the usage of any reference data. As illustrated in the example of Germany, roads in OSM are mapped completely, mainly in order of their hierarchy (Neis et al. 2012). In the beginning usually motorways are the first to be mapped completely. They are subsequently followed by municipal roads, streets in residential areas and all other roads such as forest tracks or smaller paths. Taking this information into account, a category of roads can be stated as "*close to completion*" if the monthly increase in length is very small or even close to zero. This assumption can be affirmed by a high number of active contributors along with an increasing length of mapped roads in other lower hierarchical road categories. In particular because it shows that contributors did not simply stop mapping, but instead, because of the potential completeness of a road category, switched to a lower road category which is not completely mapped yet. However, this method can only be considered as a way to approximate the quality parameter completeness. Absolute statements on the completeness of the road network are only possible with the help of a ground truth reference dataset. However, a huge benefit of this indicator can be seen in its independency of a reference dataset which makes it applicable for any region in the world.

Beside the completeness, the **logical consistency of the OSM road network** is also one key element in routing applications. In accordance with Neis et al. (2012) three topological errors are taken into account by means of internal tests: (1) roads which are erroneously not connected to each other at junctions; (2) duplicate road geometries; and (3) intersecting roads without a

common Node (3). The first inconsistency is identified by analyzing roads which do not share a common Node with another one and and lie within a radius of one meter. The second inconsistency is identified by calculating duplicate road geometries. The third inconsistency is detected by analyzing roads which intersect but do not share a common Node. This can also be caused by missing tags characterizing bridges or tunnels. These topological errors are calculated, quantified and subsequently visualized each on a single map. The relevant parameters of this section are:

- Completeness of the OSM road network
- Logical consistency of the OSM road network

3.3.2 Geocoding

The process of associating exact geographic locations with data such as street names or house numbers is generally referred to as geocoding (Amelunxen 2010) and plays a key role in many LBS applications. For this purpose, complete address information is necessary. By now, the OSM community has widely agreed on the so-called "Karlsruhe Schema" as a way to add addresses to OSM (OpenStreetMap 2013d). Within this schema house numbers are mapped either as single Nodes, as additional tags to existing features or as interpolation lines determining the start and end house number of a specified line (Ramm et al. 2011). As applications in LBS are not necessary capable of utilizing all these three methods, the **overall distribution of house numbers** over time gives a first impression of the fitness for one's needs. Other applications might be interested in the number of OSM features containing a **complete address annotation**. In comparison with the overall distribution of house numbers or house names the cases with complete annotations demonstrate the attribute completeness of the existing address information according to the aforementioned "Karlsruhe Schema". This is of particular importance for LBS, which are not able to calculate parts of an address by means of spatial queries from administrative boundaries and, furthermore, shows the attribute completeness of the appropriate features.

Moreover, good routing and navigation applications are characterized by accurate geocoding results up to the level of single buildings. To this end, all **buildings which are likely to contain a house number or house name** are calculated. Doing this, not only are actually annotated building polygons considered but also information derived from spatially intersecting Nodes or interpolation lines with address information is taken into account. By now, no algorithm is known which can distinguish between buildings which should have a house number or house name or not. Therefore, all buildings with a smaller basis than 10 m^2 and with a specified list of tags are excluded (e.g. *building = roof*, *building = garage*, etc.). Figure 2

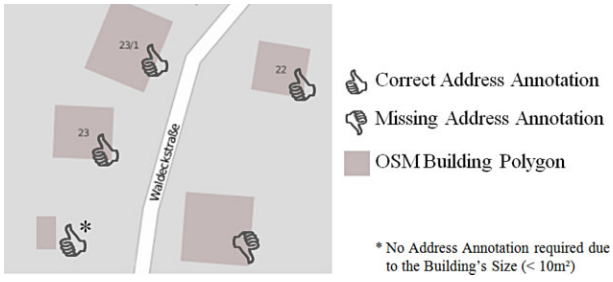

Figure 2 Buildings which are likely to contain a House Number/Name (Basemap: © OpenStreetMap contributors)

visualizes this issue. The bottom right building is erroneously not annotated with a house number/name whereas the bottom left one, due to its size being less than 10 m^2, does not need a house number/name. The latter can be presumed to be a small hut without an official house number or house name. Subsequently the development of the ratio between all buildings and those actually containing a number or name can be taken as an intrinsic indicator about the data's attribute completeness. Ideally, the cumulated number of house numbers and house names always corresponds to number of buildings, even if the number of buildings increases significantly (e.g. due to an import, a mapping party or better aerial images). The following list is a summary of the relevant parameters for geocoding:

- Buildings which should contain a house number/name
- Complete address annotation
- Overall distribution of house numbers/names

3.3.3 Points of interest-search

Points of Interests (POI) are important locations such as, for example, sights, restaurants or bus stops. In OSM, these are geographically represented by Nodes, Ways or Relations tagged with specific key-value pairs and loom large in several LBS applications. In this investigation all POIs are divided into the following nine thematic groups: accommodation and gastronomy, education, transport, finance, health care, art and culture, shop, tourism and others. Within these groups the **quantitative development of all appropriate POIs** can act as a first quantitative indicator. In general, an increasing number of POIs is a positive indicator, as the dataset is nearing completion. Beside the actual number of POIs, their detailed characterization by means of attributes has an impact on the feature's quality. In this investigation the authors hypothesize that a growing number of tags listed in the OSM map features in general increases the quality of the POIs because their characteristic values approximate reality more closely. This is especially true if the overall number of POIs is also increasing. However, it has to be stated that the development of the **average number of tags** can only act as an indicator because some OSM editors automatically assign tags to edited features which have to be filtered out.

Besides the quantitative development and the average number of attributes, the substantive differentiation within the POIs' attributes allows statements on the relative **attribute completeness** and therefore on the relative thematic accuracy. For this purpose a list of relevant keys selected from the OSM wiki is suitable (Keßler and de Groot 2013) which describe the features of the aforementioned nine groups in more detail. The list consisting of the keys (e.g. name, opening_hours, operator, website, addr:housenumber, phone, wheelchair) is adapted to the individual case, as not all POI groups necessarily need to be annotated with all of them. The development of their percentage of relative completeness is an intrinsic measurement of attribute completeness, indicating how well a respective group of POIs is suited to specified use-cases in LBSs. An advantage of this procedure is that meaningful results can be achieved even if the POIs are not completely mapped. Using predefined lists of attributes which characterize qualitative completely attributed POIs is a promising approach to assessing attribute completeness without using a reference dataset for comparison purposes. The following list contains a summary of the relevant parameters:

- Attributive Completeness of the POIs
- Average number of the POIs' tags
- Quantitative development of POIs

3.3.4 Map-applications

Beside the aforementioned use, OSM data is also widely used in map-applications. Earth surface characteristics within OSM are mainly represented by means of polygons, for instance tagged with a *natural* (e.g. glacier, wood or wetland) or a *landuse* (e.g. forest, residential area or vineyard) key. The accuracy of their geometric representation highly depends on the source (GPS traces, bulk imports or aerial images) and the acquisition scale of the contributed data (Mueller et al. 1995). A good way to determine the quality of these polygons is to calculate the equidistance between the polygons' adjacent vertices (Mooney et al. 2010b). In the proposed framework this approach is extended by the polygons' history. Comparing an initially created polygon with its currently valid one, the **evolution of the equidistance** serves as an intrinsic indicator for the relative quality development. The lower the equidistance of the currently valid polygon compared with its initially created version, the better the polygon's relative quality development, due to further editing which potentially led to a more precise geometric representation. However, several facts have to be considered: the algorithm of the tool (OSM-History-Splitter 2013, https://github.com/MaZderMind/osm-history-splitter), that is used to split OSM data into smaller extracts, has an effect on polygons lying on the boundary of the selected bounding box (especially if polygons where moved there during their history). Using the hardcut algorithm, polygons are cropped at their last Node located within the bounding box. Furthermore, divided or merged polygons can also lead to biases within the calculated equidistance because of their significantly increased or decreased area. To exclude these outliers it was chosen iteratively to consider only polygons which do not differ in size between the two compared versions by more than 50%.

The intra-theme consistency as a part of the parameter logical consistency has a major influence on the quality of a spatial dataset. This is depicted by means of **erroneously overlapping land use polygons**. Within OSM, polygons attributed with a "landuse" tag represent the primary use of an area. Basically, these polygons should not overlap each other to avoid inconsistencies possibly leading to slivers, among other reasons. These overlaps are mainly caused by inaccurate digitizing or data imports, because in each case spatial integrity of the contributions is not necessarily examined (Girres and Touya 2010). Nevertheless, sometimes a manifold land use of an area makes sense (e. g. militarily used forests). To take this fact into consideration, only overlaps with a size of less than 10% of the origin polygons are taken into account, because they more probably represent unintended overlaps. The lower the number of these detected cases, the better the intra-theme consistency concerning the land use polygons within the dataset. The following list summarizes the relevant parameters in this section:

- Erroneously overlapping land use polygons
- Evolution of the natural features' equidistance

4 Experimental Analyses and Results

In this section the results of the selected intrinsic quality indicators are outlined. For this purpose the cities of San Francisco (USA), Madrid (Spain), and Yaoundé (Cameroon) have been chosen. San Francisco is characterized by several bulk imports and a moderate-sized community. Representing a European metropolis, Madrid, as a counterpart to a US city equal in size, was chosen due to its moderate community activity without bigger imports. In contrast, the city of Yaoundé is a good example of a bulk import with no active mapping community. From all of the aforementioned indicators of the framework the following four are illustrated:

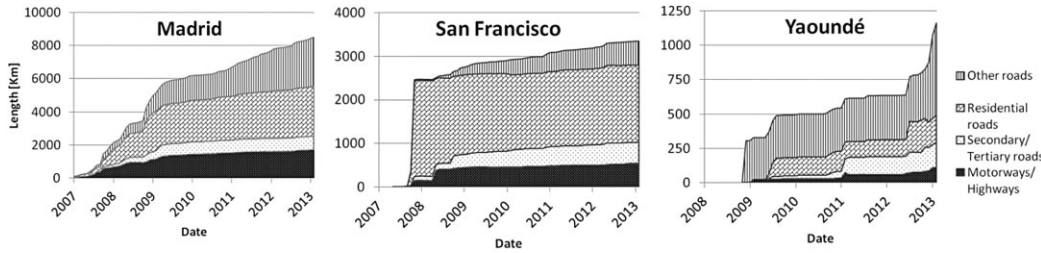

Figure 3 Development of the OSM road network length by street category for the cities of Madrid, San Francisco and Yaoundé

road network completeness, the dataset's positional accuracy, house number completeness and the geometric representation of natural polygons.

4.1 Road Network Completeness

Figure 3 shows the total road network length for the selected cities. The results clearly illustrate differences concerning their possible completeness using the abovementioned intrinsic indicators (cf. Section 3.4.1):

San Francisco shows stable lengths for motorways/highways from May 2011 until today (491 km) whereas the length of secondary/tertiary roads (480 km) and residential roads (1,790 km) does not increase significantly from April 2012. Except for the category "other roads" the road network therefore can be referred to as possibly close to completion. The strong increase of residential roads in October 2007 is accounted for by the TIGER/Line import. Madrid shows a similar pattern for secondary/tertiary roads which remain stable in length (814 km) from August 2012. All other road categories are still being mapped, although with varying intensity. Motorways/highways show an increase of approximately 10 km within the last few months whereas the categories residential roads and other roads show a much higher average amount of contribution. Minor changes in length are not necessarily new roads but can also be caused by changing the value of the highway key. In contrast, the diagram of Yaoundé reveals a stepped contribution with longer periods of no contribution at all. This suggests a hardly active community and possible data imports. Taking the small amount of active contributors into account (see Figure 4) in the case of Yaoundé, no statements on the road network completeness are possible without using a reference dataset.

4.2 Positional Accuracy of the Dataset

As described in Section 3.3, comparing actual road junctions with the previous location before their last modification gives insights into possible positional inaccuracies, for instance triggered by displaced aerial images or bulk movements. Figure 5 shows a uniform distribution of points around the centre of Madrid and San Francisco. However, the city of Madrid shows some special characteristics: in five cases two road junctions show exactly the same distance and angle to each other, indicating a possible bulk movement of the OSM data. It has to be mentioned that within this method only junctions are selected which are clearly identifiable by means of their adjacent road names. Therefore it is possible that not only the identified roads, but also other roads or even other features situated nearby could be affected by a possible bulk movement. Furthermore, within Madrid's polar scatter plot a vast amount of points has been

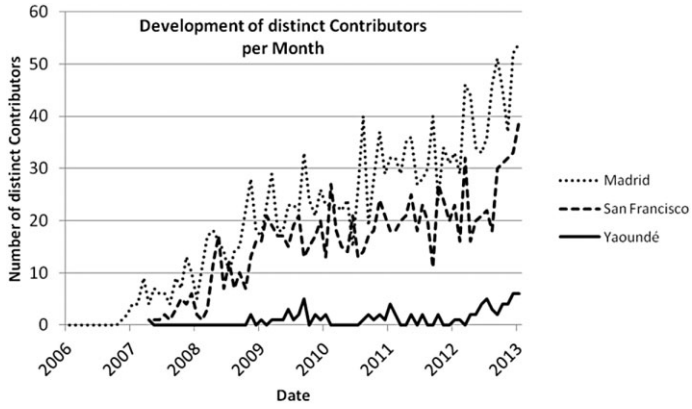

Figure 4 Number of distinct contributors per month for the cities of Madrid, San Francisco and Yaoundé

Figure 5 Polar scatter plot of degree and distance between currently visible road junctions and their previous location for the cities of Madrid, San Francisco and Yaoundé

shifted by up to 15 m between 240° and 300°. This can mean either improvement or deterioration of the positional accuracy. A precise distinction can only be made if changesets, source tags or underlying aerial images are further investigated. However, this gives a hint where further analyses are needed, potentially carried out with a ground truth reference dataset.

4.3 Buildings with a House Number/Name

In Figure 6, the cities of Madrid and Yaoundé both show a significant increase of new buildings within one month, which, due to their vast amount, is probably caused by bulk imports. This specifically applies to the city of Yaoundé where an average of 1.2 users (max: 7 users; min: 0 users) are active per month and 123,204 building polygons were imported in November 2012. In total, only four buildings are annotated with a house number/name indicating very low attribute completeness. In contrast, 1,146 (10.2%) of all buildings within San Francisco are tagged with a house number/name, whereas Madrid takes up a middle position with 1,024 (4.0%) tagged buildings. Nevertheless, in terms of attribute completeness all exemplarily investigated cities show a relatively low number of house numbers/names. Furthermore, in each of these three cases the number of created buildings does not increase proportionally.

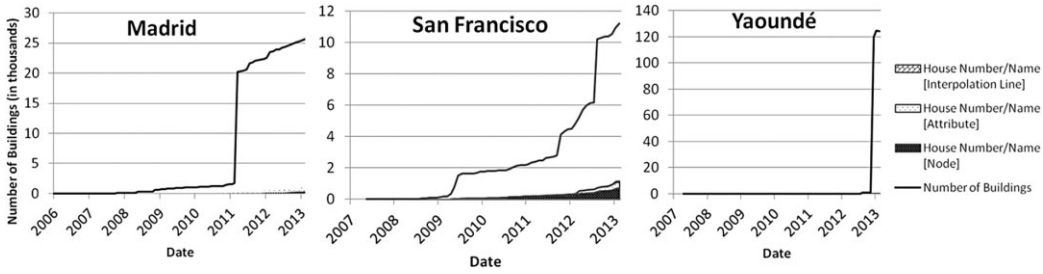

Figure 6 Development of buildings (with a House Number/Name) for the cities of Madrid, San Francisco and Yaoundé

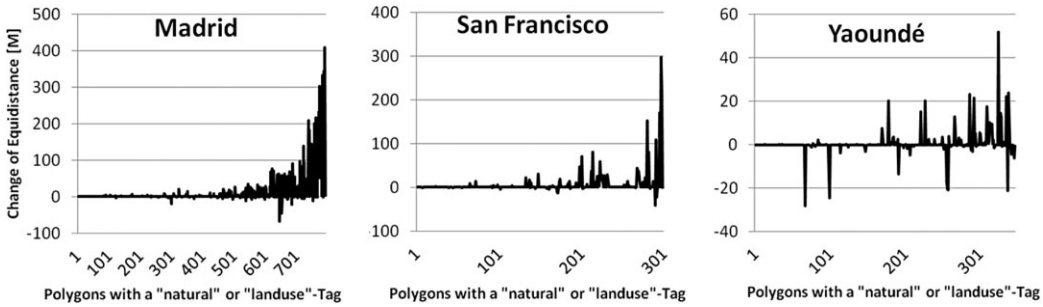

Figure 7 Equidistance development of polygons tagged with a *natural* or *landuse* tag for the cities of Madrid, San Francisco and Yaoundé

4.4 Development of Natural Polygons' Geometrical Representation

Figure 7 shows the development of the equidistance of polygons tagged with a *natural* or *landuse* tag as described in Section 3.4.4. They are sorted by the equidistance in their first version in ascending order. Examining the equidistance's development of the three selected cities several facts can be determined: with an improvement of the equidistance by an average of 11.9 m Madrid shows the highest increase (San Francisco: 6.1 m; Yaoundé: 0.5 m). Furthermore, the geometric representation of 29.5% of all investigated polygons were improved (San Francisco: 25.6%; Yaoundé: 20.1%). As Figure 7 indicates within the city of Madrid in particular polygons with a high equidistance have been improved significantly during their history. However, the majority of the polygons' geometry in all three cities has not been changed (Madrid: 62.7%; San Francisco: 67.4%; Yaoundé: 62.9%).

4.5 Architecture Framework

Figure 8 illustrates the entire architecture and workflow of the developed framework. The iOSMAnalyzer is implemented as a command line-based tool running on the Linux operating system. It is written in the Python programming language and based solely on open source components. As carefully evaluated beforehand, cropping features with the softcut algorithm (OSM-History-Splitter 2013, https://github.com/MaZderMind/osm-history-splitter) leads to distorted statistics in some cases, especially when a single version of a feature's history lies

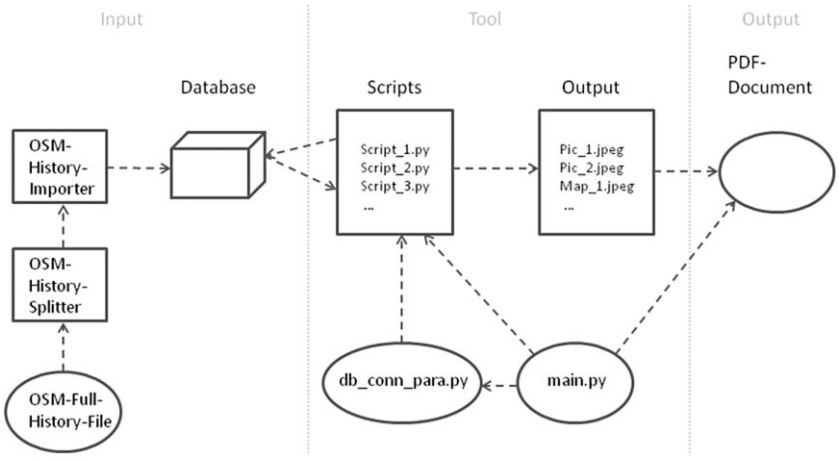

Figure 8 Architecture of the iOSMAnalyzer framework

mainly beyond the chosen bounding box. With the help of the OSM-History-Importer 2013 (https://github.com/MaZderMind/osm-history-renderer/tree/master/importer) the clipped data is imported to a PostgreSQL/PostGIS database. Resting upon this database several previously developed Python scripts compute the results from the data which are plotted to a PDF file. This file contains diagrams, tables, results of statistical analyses and maps expressing and visualizing several computed intrinsic quality indicators. Finally, it has to be mentioned that at the time of this research the OSM-History-Importer did not represent deleted Way features as such within the database. This potentially can lead to minor biases in some analyses.

5 Conclusions and Future Work

In this investigation a framework containing a broad range of more than 25 different methods and indicators is presented to evaluate the quality of an OSM dataset based on an OSM-Full-History-Dump. The holistic and thorough intrinsic approach carried out in this investigation allows data quality evaluations without a ground truth reference dataset. This is beneficial in many respects: accessibility to high quality and commercial datasets is often limited due to high costs and contradictory licensing restrictions. These facts allow OSM quality analyses without any regional or financial limitations. The calculated results are provided in the form of statistics, tables, diagrams and maps and give a compact quality overview of a freely selectable area. As quality heavily depends on the individual use case, the OSM data is evaluated in terms of "Fitness for Purpose" for LBSs concerning the categories "General Information on the Study Area", "Routing and Navigation", "Geocoding", "Points of Interest-Search", "Map Applications" and "User Information and Behavior". However, absolute statements on data quality are only possible with a high quality reference dataset as a basis for comparison. Nevertheless, in an intrinsic approach quality parameters, as for example the road network completeness can only be determined approximately; in this example by investigating the historical increase of mapped roads within different road categories. Furthermore, the contributor activity also has an effect on OSM data. This investigation revealed that the interpretation of some intrinsic quality indicators is facilitated and supported by means of contributor activity.

For future work Relations (e.g. turn restrictions, bus routes, etc.) should be taken into consideration. Currently, the OSM-History-Importer does not support the import of Relations to the database, therefore these were not considered in this research. Furthermore, as the quality of OSM data also depends on the project's contributors, more in-depth analyses regarding their experience, quality of contributions and reputation have to be integrated into the framework. Moreover some of the proposed methods could be evaluated by means of a ground truth reference dataset. The higher the conformity of the intrinsic results within this comparison, the better the proposed indicator is possibly suited. Additionally, a pre-calculated signature database containing different patterns of quality manifestations could serve as a reference for other OSM areas with similar characteristics (e.g. community activity, size or spatial structure). Due to its modular structure, the implemented framework can easily be extended by further methods and indicators.

References

Amelunxen C 2010 On the suitability of Volunteered Geographic Information for the purpose of geocoding. In *Proceedings of the Geoinformatics Forum*, Salzburg, Austria

Batini C and Scannapieco M 2006 *Data Quality: Concepts, Methodologies and Techniques*. Berlin, Springer

Bing 2010 Bing Engages Open Maps Community. WWW document, http://www.bing.com/blogs/site_blogs/b/maps/archive/2010/11/23/bing-engages-open-maps-community.aspx

Brando C and Bucher B 2010 Quality in user generated spatial content: A matter of specifications. In *Proceedings of the Thirteenth AGILE International Conference on Geographic Information Science*, Guimarães, Portugal

Budhathoki N R 2010 Participants' Motivations to Contribute Geographic Information in an Online Community. Unpublished Ph.D. Dissertation, University of Illinois, Urbana-Champaign

Chilton S 2009 Crowdsourcing is radically changing the geodata landscape: Case study of OpenStreetMap. In *Proceedings of the Twenty-fourth International Cartography Conference*, Santiago, Chile

Ciepłuch B, Mooney P and Winstanley A 2011 Building generic quality indicators for OpenStreetMap. In *Proceedings of the Nineteenth Annual GISRUK Conference*, Portsmouth, England

Coleman D J, Georgiadou Y, and Labonte J 2009 Volunteered geographic information: The nature and motivation of producers. *International Journal of Spatial Data Infrastructures Research* 4: 332–58

Devillers R, Bédard Y, Jeansoulin R, and Moulin B 2007 Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Science* 21: 261–82

Devillers R, Gervais M, Bédard Y, and Jeansoulin R 2002 Spatial data quality: From metadata to quality indicators and contextual end-user manual. In *Proceedings of the OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management*, Istanbul, Turkey

van Exel M, Dias E, and Fruijtier S 2010 The impact of crowdsourcing on spatial data quality indicators. In *Proceedings of the Sixth International Conference on Geographic Information Science (GIScience 2010) Workshop on the Role of Volunteered Geographic Information in Advancing Science*, Zurich, Switzerland

Flanagin A J and Metzger M J 2008 The credibility of volunteered geographic information. *GeoJournal* 72: 137–48

Girres J F 2011 A model to estimate length measurements uncertainty in vector databases. In *Proceedings of the Seventh International Symposium on Spatial Data Quality (ISSDQ'11)*, Coimbra, Portugal

Girres J-F and Touya G 2010 Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS* 14: 435–59

Goetz M and Zipf A 2013 The evolution of geo-crowdsourcing: bringing volunteered geographic information to the third dimension. In Sui D, Elwood S and Goodchild M F (eds) *Crowdsourcing Geographic Knowledge*. Berlin, Springer: 9–59

Goodchild M F 1995 Sharing imperfect data. In Onsrud H J and Rushton G (eds) *Sharing Geographic Information*. New Brunswick, NJ, Rutgers University Press: 413–25

Goodchild M F 2007 Citizens as sensors: The world of volunteered geography. *GeoJournal* 69: 211–21

Goodchild M F 2009 NeoGeography and the nature of geographic expertise. *Journal of Location Based Services* 3: 82–96

Goodchild M F and Li L 2012 Assuring the quality of volunteered geographic information. *Spatial Statistics* 1: 110–20

Hagenauer J and Helbich M 2012 Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks. *International Journal of Geographical Information Science* 26: 963–82

Haklay M 2010 How good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B* 37: 682–703

Haklay M, Basiouka S, Antoniou V, and Ather A 2010 How many volunteers does it take to map an area well? The validity of Linus' law to Volunteered Geographic Information. *Cartographic Journal* 47: 315–22

Helbich M, Amelunxen C, Neis P, and Zipf A 2012 Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata. In Strobl J, Blaschke T, and Griesebner G (eds) *Angewandte Geoinformatik 2012*. Berlin, Herbert Wichmann Verlag: 24–33

Hristova D, Quattrone G, Mashhadi A, and Capra L 2013 The life of the party: Impact of social mapping in OpenStreetMap. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*, Boston, Massachusetts

Keßler C and de Groot R T A 2013 Trust as a proxy measure for the quality of Volunteered Geographic Information in the case of OpenStreetMap. In Groot R T A de, Bucher B, and Crompvoets J (eds) *Proceedings of the Sixteenth AGILE Conference on Geographic Information Science*. Berlin, Springer Lecture Notes in Geoinformation and Cartography: 21–37

Keßler C, Trame J, and Kauppinen T 2011 Tracking editing processes in Volunteered Geographic Information: The case of OpenStreetMap. In *Proceedings of the Conference on Spatial Information Theory (COSIT '11) Identifying Objects, Processes and Events in Spatio-Temporally Distributed Data (IOPE) Workshop*, Belfast, Maine

Kounady O 2011 Assessing the quality of OpenStreetMap data. Unpublished M.Sc. Thesis, Department of Civil, Environmental and Geomatic Engineering, University College of London

de Leeuw J, Said M, Ortegah L, Nagda S, Georgiadou Y, and DeBlois M 2011 An assessment of the accuracy of volunteered road map production in western Kenya. *Remote Sensing* 3: 247–56

Lin Y-W 2011 A qualitative enquiry into OpenStreetMap making. *New Review of Hypermedia and Multimedia* 17: 53–71

Ludwig I, Voss A, and Krause-Traudes M 2011 A comparison of the street networks of Navteq and OSM in Germany. In Geertman S, Reinhardt W, and Toppen F (ed) *Advancing Geoinformation Science for a Changing World*. Berlin, Springer: 65–84

Mondzech J and Sester M 2011 Quality analysis of OpenStreetMap data based on application needs. *Cartographica* 46: 115–25

Mooney P and Corcoran P 2012a Characteristics of heavily edited objects in OpenStreetMap. *Future Internet* 4: 285–305

Mooney P and Corcoran P 2012b The annotation process in OpenStreetMap. *Transactions in GIS* 16: 561–79

Mooney P, Corcoran P, and Winstanley A 2010a Towards quality metrics for OpenStreetMap. In *Proceedings of the Eighteenth ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, San Jose, California: 514–17

Mooney P, Corcoran P, and Winstanley A 2010b A study of data representation of natural features in OpenStreetMap. In *Proceedings of the Sixth International Conference on Geographic Information Science*, Zurich, Switzerland

Mueller J-C, Lagrange J-P, and Weibel R 1995 *GIS and Generalization: Methodology and Practice*. London, Taylor and Francis

Napolitano M and Mooney P 2012 MVP OSM: A Tool to Identify Areas of High Quality Contributor Activity in OpenStreetMap. WWW document, https://github.com/napo/mvp-osm

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

Neis P, Zielstra D, and Zipf A 2013 Comparison of volunteered geographic information data contributions and community development for selected world regions. *Future Internet* 5: 282–300

Neis P and Zipf A 2012 Analyzing the contributor activity of a Volunteered Geographic Information project: The case of OpenStreetMap. *ISPRS International Journal of Geo-Information* 1: 146–65

Nielsen J 2006 Participation Inequality: Encouraging More Users to Contribute. WWW document, http://www.nngroup.com/articles/participation-inequality/

O'Reilly T 2005 What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. WWW document, http://oreilly.com/web2/archive/what-is-web-20.html

OpenStreetMap 2013a OpenStreetMap Import Catalogue. WWW document, http://wiki.openstreetmap.org/wiki/Import/Catalogue

OpenStreetMap 2013b OpenStreetMap Map Features. WWW document, http://wiki.openstreetmap.org/wiki/Map_Features

OpenStreetMap 2013c OpenStreetMap Planet.osm/full. WWW document, http://wiki.openstreetmap.org/wiki/Planet.osm/full

OpenStreetMap 2013d OpenStreetMap Karlsruhe Schema. WWW document, http://wiki.openstreetmap.org/wiki/Proposed_features/House_numbers/Karlsruhe_Schema

OpenStreetMap 2013e OpenStreetMap Registered Users. WWW document, http://wiki.openstreetmap.org/wiki/Stats#Registered_users

Ramm F, Topf J, and Chilton S 2011 *OpenStreetMap: Using and Enhancing the Free Map of the World*. Cambridge, UK, UIT Cambridge

Rehrl K, Groechenig S, Hochmair H, Leitinger S, Steinmann R, and Wagner A 2013 A conceptual model for analyzing contribution patterns in the context of VGI. In Krisp J M (ed) *Progress in Location-Based Services*. Berlin, Springer: 373–88

Touya G and Brando-Escobar C 2013 Detecting level-of-detail inconsistencies in volunteered geographic information data sets. *Cartographica* 48: 134–43

Veregin H 1999 Data quality parameters. In Longley P A, Goodchild M F, Maguire D J, and Rhind D W (eds) *Geographical Information Systems: Principles and Technical Issues*. New York, John Wiley and Sons: 177–89

Wang R and Strong D 1996 Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems* 12: 5–33

Zielstra D and Hochmair H 2012 Comparing shortest paths lengths of free and proprietary data for effective pedestrian routing in street networks. *Transportation Research Record* 2299: 41–7

Zielstra D, Hochmair H, and Neis P 2013 Assessing the effect of data imports on the completeness of OpenStreetMap: A United States case study. *Transactions in GIS* 17: 315–34

Zielstra D and Zipf A 2010 A comparative study of proprietary geodata and volunteered geographic information for Germany. In *Proceedings of the Thirteenth AGILE International Conference on Geographic Information Science*, Guimarães, Portugal

