Lecture Notes in Geoinformation and Cartography

LNG&C

Jamal Jokar Arsanjani
Alexander Zipf
Peter Mooney
Marco Helbich *Editors*

OpenStreetMap in GIScience

Experiences, Research, and Applications

Springer

Lecture Notes in Geoinformation and Cartography

About the Series

The Lecture Notes in Geoinformation and Cartography series provides a contemporary view of current research and development in Geoinformation and Cartography, including GIS and Geographic Information Science. Publications with associated electronic media examine areas of development and current technology. Editors from multiple continents, in association with national and international organizations and societies bring together the most comprehensive forum for Geoinformation and Cartography.

The scope of Lecture Notes in Geoinformation and Cartography spans the range of interdisciplinary topics in a variety of research and application fields. The type of material published traditionally includes:

- proceedings that are peer-reviewed and published in association with a conference;
- post-proceedings consisting of thoroughly revised final papers; and
- research monographs that may be based on individual research projects.

The Lecture Notes in Geoinformation and Cartography series also includes various other publications, including:

- tutorials or collections of lectures for advanced courses;
- contemporary surveys that offer an objective summary of a current topic of interest; and
- emerging areas of research directed at a broad community of practitioners.

More information about this series at http://www.springer.com/series/7418

Jamal Jokar Arsanjani
Alexander Zipf
Peter Mooney
Marco Helbich
Editors

OpenStreetMap in GIScience

Experiences, Research, and Applications

Springer

Editors
Jamal Jokar Arsanjani
GIScience Research Group
Institute of Geography
Heidelberg University
Heidelberg
Germany

Alexander Zipf
GIScience Research Group
Institute of Geography
Heidelberg University
Heidelberg
Germany

Peter Mooney
Department of Computer Science
Maynooth University
Maynooth, Co. Kildare
Ireland

Marco Helbich
Department of Human Geography and Spatial Planning
Utrecht University
Utrecht
The Netherlands

ISSN 1863-2246 ISSN 1863-2351 (electronic)
Lecture Notes in Geoinformation and Cartography
ISBN 978-3-319-14279-1 ISBN 978-3-319-14280-7 (eBook)
DOI 10.1007/978-3-319-14280-7

Library of Congress Control Number: 2014958977

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Foreword

OpenStreetMap Studies and Volunteered Geographical Information

This book comes at an apt time to reflect on the growing role of OpenStreetMap (OSM) in Geographical Information Science. This summer, the OpenStreetMap project celebrated ten years of operation, which began on the date of the domain name registration. I first heard about the project when it was in its very early stages and, with the support of the Royal Geographical Society, carried out the first research project that focused on OpenStreetMap, with an attempt to develop a mobile data collection tool on an early GPS-enabled phone. As a result, I found myself writing, together with Patrick Weber, what is now the most cited paper on the project (Haklay and Weber 2008). This early exposure to the project provided me with opportunities to watch, with astonishment, how it has become an important source of geographical information, as well as the explosive growth in academic research with and about it.

Of course, in the early years the project was small, with an unclear future and too localised to have a wider impact. It is, therefore, unsurprising that, so far as academic publications indexing reveals, Nelson et al. (2006) 'Towards development of a high quality public domain global roads database' and Taylor and Caquard (2006) 'Cybercartography: Maps and Mapping in the Information Era' are the first peer-reviewed papers that mention OpenStreetMap. Yet, it is interesting that, within two years of establishment, researchers in Canada and the United States heard about it and realised its potential. Moreover, many chapters in the current volume attest to the foresight that these two papers demonstrated.

Since 2006, OpenStreetMap has received plenty of academic attention. As of August 2014, more 'conservative' academic search engines such as ScienceDirect or Scopus find 286 and 236 peer-reviewed papers (respectively) that mention the project. The ACM digital library finds 461 papers in the areas that are relevant to computing and electronics, while Microsoft Academic Research finds only 112. Google Scholar, probably the most expansive of the search engines, lists over

9000 (!). Even with the most conservative version from Microsoft, we can see an impact on fields ranging from social science to engineering and physics. In short, OpenStreetMap has facilitated major contributions to knowledge beyond producing maps.

The link between OpenStreetMap and the concept of Volunteered Geographical Information is also long-standing. Michael Goodchild, in his seminal paper from 2007 that defined Volunteered Geographic Information (VGI), mentioned OpenStreetMap as an example. Since then the literature frequently conflates OSM and VGI. In some recent papers statements such as 'OpenStreetMap is considered as one of the most successful and popular VGI projects' or 'the most prominent VGI project OpenStreetMap' are common[1] and, to some degree, the boundary between the two is being blurred. I also admit to be part of the problem—for example, with the title of my 2010 paper 'How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets'. However, upon reflection on the characteristics of OpenStreetMap and other VGI projects, I became uncomfortable with the equivalence between OSM and VGI. The stance that Neis and Zielstra (2014) offer is, I suggest, more accurate: 'One of the most utilized, analyzed and cited VGI-platforms, with an increasing popularity over the past few years, is OpenStreetMap (OSM).'

The reason that it is valuable to differentiate between focusing on the OpenStreetMap project (what we can call OSM studies) and the more generic VGI research is partly due to the volume of papers specifically about the project, and what they reveal about the project. Over the years, several types of research papers that can be classified as OSM studies have emerged.

First, there is a whole set of research projects that use OSM data because it is easy to use and free to access (for example, in computer vision or even string theory). For these projects, OSM is just data to be used (see "Data Retrieval for Small Spatial Regions in OpenStreetMap" and "The Next Generation of Navigational Services Using OpenStreetMap Data: The Integration of Augmented Reality and Graph Databases", which arguably fall into this category). Second, there are studies of OSM data: quality, the history and evolution of objects in the database, what we can learn about the nature of the data and other aspects. The majority of this volume falls under this category (see "Assessment of Logical Consistency in OpenStreetMap Based on the Spatial Similarity Concept"–"Inferring the Scale of OpenStreetMap Features", "Route Choice Analysis of Urban Cycling Behaviors Using OpenStreetMap: Evidence from a British Urban Environment", "Building a Multimodal Urban Network Model Using OpenStreetMap Data for the Analysis of Sustainable Accessibility"–"Using Crowd-Sourced Data to Quantify the Complex Urban Fabric—OpenStreetMap and the Urban–Rural Index"). Third, there are studies that also look at the interactions between patterns of contribution and the data—for example, in trying to infer trustworthiness (see "Spatial Collaboration Networks of OpenStreetMap"). Fourth, there are studies that look at the wider

[1] These are deliberately unreferenced so as not to argue that specific authors are to blame.

societal aspects of OpenStreetMap—for example, what the spatial and social implications of data coverage are (see "Social and Political Dimensions of the OpenStreetMap Project: Towards a Critical Geographical Research Agenda"). Finally, there are studies of the social practices in OpenStreetMap as a project (see "The Impact of Society on Volunteered Geographic Information: The Case of OpenStreetMap").

In short, there is a significant body of knowledge regarding the nature of the project, the implications of what it produces, and ways to understand the information that emerges from it. Clearly, we now know that OSM produces good data and is aware of the patterns of contribution. What is also clear is that many of these patterns are specific to OSM. Because of the importance of OSM to so many application areas (including illustrative maps in string theory!), these insights are very important. Some of these insights are expected to be also present in other VGI projects but making such analogy needs to be done carefully, and only when there is evidence from other projects that this is the case. In short, we should avoid conflating VGI and OSM—and this volume provides a clear demonstration why this is the case.

November 2014

Prof. Mordechai (Muki) Haklay
Professor of Geographical Information Science
Department of Civil, Environment and Geomatic Engineering
University College London (UCL), UK

References

Goodchild M (2007) Citizens as sensors: the world of volunteered geography. GeoJournal 69(4):211–221

Haklay M, Weber P (2008) OpenStreetMap: user-generated street maps. IEEE Perv Comput 7(4):12–18

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

Neis P, Zielstra D (2014) Recent developments and future trends in volunteered geographic information research: the case of OpenStreetMap. Future Int 6, 1:76–106

Nelson A, de Sherbinin A, Pozzi F (2006) Towards development of a high quality public domain global roads database. Dat Sci J, 5:223–265

Taylor DRF, Caquard S (2006) Cybercartography: maps and mapping in the information era. Cartographica 41(1):1–6

Contents

Part II Social Context

Part III Network Modeling and Routing

Part IV Land Management and Urban Form

An Introduction to OpenStreetMap in Geographic Information Science: Experiences, Research, and Applications

Jamal Jokar Arsanjani, Alexander Zipf, Peter Mooney and Marco Helbich

Abstract Recent years have seen new ways of collecting geographic information via the crowd rather than organizations. OpenStreetMap (OSM) is a prime example of this approach and has brought free access to a wealth of geographic information —for many parts of the world, for the first time. The strong growth in the last few years made more and more people consider it as a potential alternative to commercial or authoritative data. The increasing availability of ever-richer data sets of freely available geographic information led to strong interest of researchers and practitioners in the usability of this data—both its limitations and potential. Both the unconventional way the data is being produced as well as its richness and heterogeneity have led to a range of different research questions on how we can assess, mine, enrich, or just use this data in different domains and for a wide range of applications. While this book cannot present all types of research around OpenStreetMap or even the broader category of User Generated Content (UGC) or Volunteered Geographic Information (VGI), it attempts to provide an overview of the current state of the art by presenting some typical and recent examples of work in GIScience on OSM. This chapter provides an introduction to the scholarly work on OpenStreetMap and its current state and summarizes the contributions to this book.

J. Jokar Arsanjani (✉) · A. Zipf
GIScience Research Group, Institute of Geography, Heidelberg University,
69120 Heidelberg, Germany
e-mail: jokar.arsanjani@geog.uni-heidelberg.de

A. Zipf
e-mail: zipf@uni-heidelberg.de

P. Mooney
Department of Computer Science, Maynooth University,
Maynooth, Co. Kildare, Ireland
e-mail: peter.mooney@nuim.ie

M. Helbich
Department of Human Geography and Spatial Planning, Utrecht University,
Heidelberglaan 2, 3584 CS Utrecht, The Netherlands

J. Jokar Arsanjani et al. (eds.), *OpenStreetMap in GIScience*,
Lecture Notes in Geoinformation and Cartography,
DOI 10.1007/978-3-319-14280-7_1

1 Introduction

Access to spatial data and cartographic products has changed radically over the last decade or so. Traditionally, governmental agencies, cartographic centers, and commercial agencies were the only sources for end-users seeking spatial data. One of the most formidable barriers to more widespread access to these geodata were created by often prohibitive high fees and license charges in combination with time- and purpose-limited copyright restrictions imposed. This business model was rather successful, but made access to high-quality geodata very difficult for all but a small number of end users. Changes in Information and Communication Technology (ICT) brought about by the Internet and social media and the availability of inexpensive portable satellite navigation devices has seen this traditional geodata business model challenged. One of the key driving forces in this change has been the OpenStreetMap (OSM) project. OSM was launched in 2004 with the mission of creating an editable map of the whole world and released with an open content license (http://wiki.openstreetmap.org/wiki/About). In general, OSM aims at building and maintaining a free editable map database of the world in a collaborative manner so that people and end-users are not forced to buy geodata in the traditional way and subsequently be subjected to restrictive copyright and license commitments. OSM started initially with a focus on mapping streets and roads. Since then it has moved far beyond these entities and it now contains a very rich variety of geographical objects (e.g., buildings, land use, Points of Interest) from all over the planet being mapped by thousands of volunteer contributors to the project. Aside from the obvious commercial benefits offered by OSM, the project has revolutionized the way in which geodata is collected. No longer are the collection of geodata and the development of cartographic products limited to specialists, geographic surveyors, or cartographers.

OSM is often referred to as the Wikipedia map of the world. As it is built on many of the same ICT structures as Wikipedia it offers its project contributors the possibility of (a) almost immediate updating of the map database as well as very frequent updating of associated editing software and other tools; (b) importing geodata recorded from Global Positioning System (GPS)-enabled devices, smartphones, and other digital maps tools; (c) access to the full history of mapping activities in OSM over its lifetime; and finally (d) collaboration with other OSM users and contributors through various communication channels including mailing lists, discussion forums, and physical meetings (Mooney and Corcoran 2013a). The gradual evolution of the OSM ecosystem has been very successful. The project got off to a slow start but since 2007 there has been an ever-increasing rate of people joining the project. In November 2014, OSM had approximately 1.85 million registered users and contributors (http://wiki.openstreetmap.org/wiki/Stats). As mentioned previously, the era of ubiquitous Internet, social media, open-source software, etc. has seen many citizen knowledge-based projects for a host of diverse purposes launched on the Internet over the last few years. OSM has been a unique case. The academic and industrial communities have recognized OSM not solely based on its rise to become an important distributor of geodata but its wider success in growing a global community of people willing to participate in the collection and maintenance of geodata. The OSM community is actively involved in much more than collecting geodata to build and maintain this global geodatabase. In addition, the community is involved in, for example, humanitarian work, open source software development to support OSM and the GIS community, and in building a network of support for those using and contributing to the OSM project.

In recent years, several scientific disciplines (e.g. geography, GIScience, spatial planning, cartography, computer science, and ecology) have realized the immense potential of OSM and it has become the subject of academic research. OSM offers researchers a unique dataset that is global in scale and a body of knowledge created and maintained by a very large collaborative network of volunteers. Research on OSM has shown that its geodata in some parts of the world are more complete and locationally and semantically more accurate than the corresponding proprietary datasets (e.g., Zielstra and Zipf 2010; Neis and Zipf 2012; Helbich et al. 2012), while being of high spatial heterogeneity. Skepticism amongst the GIS community and industry surrounding the quality of the geodata in OSM has seen a major effort being made on evaluating the quality of the OSM geodata. This has a led to the development of a number of software tools and methodologies for analyzing the quality (Roick et al. 2011; Helbich et al. 2012; Jokar Arsanjani et al. 2013a; Jokar Arsanjani and Vaz 2015c). Other approaches even try to improve the OSM data through algorithms dedicated to specific object types, such as addresses for geocoding (Amelunxen 2010). Investigation of the development and evolution of OSM across the globe over time has also emerged as a research topic for many academic studies (Mooney et al. 2012; Neis and Zipf 2012; Jokar Arsanjani et al. 2013c; Mooney and Corcoran 2013; Fan et al. 2014).

Extracting value-added information from the OSM database has become another emerging research topic for researchers to attempt to understand OSM better (Hagenauer and Helbich 2012; Mooney and Corcoran 2012; Mooney et al. 2013; Jokar Arsanjani et al. 2015). Hagenauer and Helbich (2012), for instance, predicted missing urban areas through artificial neural networks. Bakillah et al. (2014a) derived population estimations from OSM and an emerging important topic is land use maps that can be generated using OSM (Jokar Arsanjani et al. 2014; Jokar Arsanjani and Vaz 2015c). Klonner et al. (2014) investigated the updating of Digital Elevation Models and Fan et al. (2014) estimated building types from OSM.

Both inside and outside of the academic sphere, OSM is now being used increasingly in a variety of practical or scientific applications in different domains, which demonstrates the usability of the crowdsourced geodata in OSM. However, in all of these cases the characteristics of OSM must be considered. Because of the flexibility and open data-like structure of OSM, it is possible to use or even adapt and improve OSM for a large range of purpose-directed applications, as we will see below. As mentioned above, there are some data quality issues with the OSM database which can be mitigated against through specialized approaches to using the actual geodata (Goodchild 2013). This has brought about a host of examples of applications and domain-specific research. A first important category is the development of a set of different special routing and navigation systems that operate on a large scale. Examples include: routing for cars, bikes, and pedestrians such as in OpenRouteService, (Schmitz et al. 2008); emergency routing (Neis et al. 2010); wheelchair routing (Neis 2014); emergency response and evacuation simulation (Bakillah et al. 2012); indoor routing (Goetz 2012); or agricultural logistics (Lauer et al. 2014). Further typical uses of OSM include improving cartography (Rylov and Reimer 2014) or developing Location Based Services (LBS) (Schilling et al. 2009). Another innovation was the development of 3D city models from OSM (Over et al. 2010; Goetz 2013). Further research has focused on attempting to extend the current OSM spatial data model by working on extensions such as: 3D (Goetz and Zipf 2012), indoor mapping (Goetz and Zipf 2012), or wheelchair routing (Neis 2014) and using the results from this in a range of applications.

The relationship between OSM and open data standards, in particular Spatial Data Infrastructures (SDI) and the future direction of the Web 2.0 paradigm, is a question still requiring further discussion. In particular, the large volumes of data being updated by the minute that are now available pose challenges with regards to their handling and keeping them up to date on a global scale.

The discussion in the preceding paragraphs has shown that OSM has now emerged as a new research area. It has the potential to bring disparate research disciplines together and enhances interdisciplinary and multidisciplinary investigations. This interdisciplinary research collaboration can contribute to a more profound and cross-disciplinary understanding of citizens' knowledge-based efforts in projects such as OSM. It also provides an interesting platform for the academic research community to collaborate with these communities towards interactive collection of up-to-date geodata from citizens by means of novel computationally oriented methods such as network analysis, machine learning, and computer simulation models. As the examples above have demonstrated, these practical investigations on OSM provide a rich set of opportunities to discover novel and valuable patterns inherent in the geodata collected by citizens, to better understand the activities of contributors to open knowledge projects, the characteristics of their human–computer interactions, and the potential to tackle classical GIS research questions using this modern and revolutionary approach to the collection and distribution of geographic data.

2 A Short Overview of the OpenStreetMap Research Landscape

In this section, we present a brief overview of the OSM research landscape through a word cloud approach. To do so, a search query was applied on 16 August 2014 in Google Scholar looking for four terms "OpenStreetMap", "OSM", "VGI", and "Volunteered Geographic Information" either in the abstracts, titles or keywords. In total, 224 documents were collected. The collection of titles, abstracts, and keywords were explored by means of word clouds. Word clouds provide an intuitive

Fig. 1 Word cloud of the papers' titles

impression about common words and show the number of times a certain word appeared in the literature. This is expressed by varying font sizes. Larger font sizes refer to words that appear more often than smaller font sizes (Helbich et al. 2013). It should be noted that the aforementioned search terms were removed from the resulting word cloud as their usage frequencies were substantial and masked the other terms. We leave these figures for the readers to interpret them visually and gain some insights about the research on OSM so far.

As editors of this volume, we have been involved in research connected to OpenStreetMap for many years. From our own empirical experience, these word

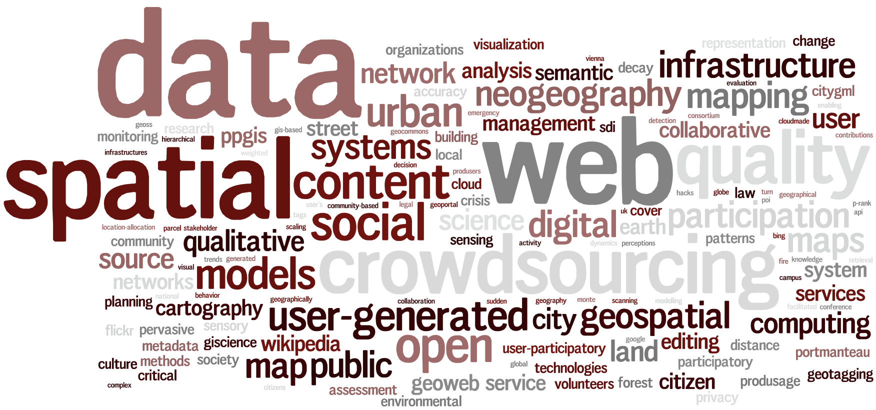

Fig. 2 Word cloud of the papers' keywords

Fig. 3 Word cloud of the papers' abstracts

clouds capture the essence of academic research on OpenStreetMap from the past number of years. In Figs. 1 and 2 where the word clouds of paper title terms and keywords are presented we see a few dominant terms, particularly *data*, *quality*, *research*, *systems,* and *participation*. As we mentioned earlier, the issue of the quality of OSM data has been on the research agenda for many years now. It is likely to remain on the agenda for some time to come. However, these titles also reflect the expansion of OSM research to consider the citizen and volunteer participation, which drives the expansion of the project. The systems, models, sensors, and applications which collect, analyze, store, manage, transform, and distribute the data must now be studied and explored in more detail as OpenStreetMap grows in size potentially towards being considered as Geographic Big Data (Goodchild 2013).

The word cloud in Fig. 3 displays the most frequently occurring terms in the abstracts of the papers returned in our Google Scholar searches. Quality is a dominant term but we see the concepts of research, social, community, methods, and development emerging. There are obvious visual linkages to the word cloud in Fig. 4 where the abstracts of the chapters included in this volume are visualized.

Fig. 4 Word cloud of our chapters' abstracts

We can immediately see the same set of dominant terms. However, in our abstracts word cloud there is somewhat more diversity with urban modelling, navigation, modelling, and knowledge management related terms being highlighted.

3 Geography of OpenStreetMap

As already stated in the literature (e.g., Mooney and Corcoran 2012; Neis 2014), OpenStreetMap has its own geography across time and space. In other words, we rarely see identical patterns of contributions in two different regions/countries. When speaking of OpenStreetMap quality and contributions networks the importance of studying diverse case studies has been highlighted. Hence, in this section, two different maps are generated from the OSM statistics, which demonstrate the heterogeneity of OSM in different countries. Figure 5 displays the total number of created nodes in October 2014. This map displays a thematic categorization of created nodes, which is one of the key elements in measuring OSM contributions. It should be noted that in this comparative report, the size of the country, population, gross domestic product (GDP), and a number of other physical characteristics of the countries are not taken into consideration. However, they are of great importance in performing further in-depth analysis. For instance, the dominant land cover types in Canada and Australia should be excluded in considering the size of the country as apart from land cover there are no objects to be mapped and the contributed nodes have very likely occurred within urban areas. Besides, their populations are not comparable to the USA, China, and India.

Nonetheless, focusing on count gives an overall indication that the high number of node creation is not limited to European countries, but other countries are also

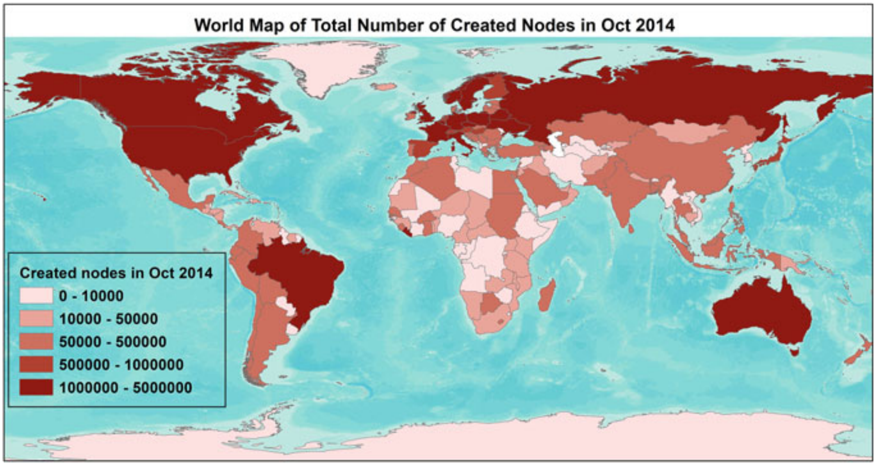

Fig. 5 A world map of the total created nodes in October 2014

emerging in OSM. Amongst these emerging countries, north America including USA and Canada, south American countries particularly Brazil, Australia, and some Asian and African countries can be named, which calls for further studies in these regions to find out how actively and accurately mapping in OSM is being undertaken. In terms of number of created nodes, in total over 46 million nodes were created in this month. In a number of countries, no nodes were created. However, Germany, United States, Russia, Czech Republic, Italy, Poland, France, Norway, Liberia, Canada, and Japan received the most created nodes, respectively.

In terms of total active members, i.e., mappers, in this period, while in total 3,048 members logged into OSM, the majority of them were from Germany (535), United States (215), Russia (212), France (195), Italy (156), Poland (155), UK (128), Spain (96), Austria (81), and Japan (55). In order to normalize the number of active members, the average number of active members per day in October 2014 is divided by the total population of the countries in terms of millions of people in 2010. Figure 6 shows the average number of active members per day for October 2014 per million people. This map helps to detect the countries that have a large portion of their population involved in the mapping process in OSM. Italy, the Netherlands, Kuwait, Croatia, and Liberia were at the top of the list. It is interesting to see that a number of countries from all continents have more than 1 member per million population active in mapping. On the contrary, a number of Asian and African countries have a very minor proportion of their population involved in mapping. This confirms the empirical findings that only a small portion of the population is mapping. It is worth mentioning that this finding is based on our analysis within the chosen timeframe for sharing general impressions about OSM and activities in OSM certainly also has a temporal pattern, which is an important indicator to be considered.

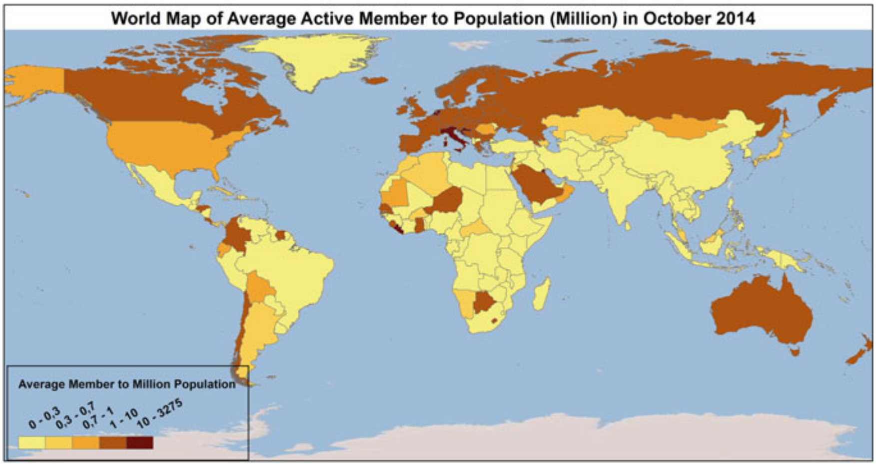

Fig. 6 A world map of average number of active members to population (million) in October 2014

It is interesting to note that contrary to Germany as an active country in OSM, some other countries are emerging in OSM and there are still large gaps in the OSM data from these countries. However, this can be viewed positively. Information dissemination about OSM within the last few years has promoted additional people to become members of OSM. Slowly but surely OSM is gaining popularity in these countries. Perhaps OpenStreetMap is helping address the participation inequality that is strongly represented in many types of User-Generated Content on the Internet today. This "Digital Divide" indicates that very small groups with specific demographic and geographical characteristics are responsible for production of most of the UGC we see on the Internet today. However, these map visualizations indicate that OpenStreetMap is reaching into countries and regions which heretofore would have felt the consequences of the digital divide (Graham et al. 2012). Improvements in ICT infrastructures and IT education for socially deprived groups such as women and children coupled with more ubiquitous access to smartphone technology has provided an environment where participation in OpenStreetMap can increase. Research will need to be undertaken to gain a better understanding of the social processes involved in these changes.

4 Objectives and Scope

The present volume entitled "OpenStreetMap in Geographic Information Science: experiences, research, and applications" presents a collection of experiences and research which has been carried out with OSM as the central and core theme. The book seeks to build a firm foundation for research work focused on integrating OSM. Chapters will address the following research topics:

(a) State-of-the-art and cutting-edge approaches for data quality analysis in OSM and VGI.
(b) Investigations on understanding OSM contributors and the nature of their contributions.
(c) Identification of patterns of contributions and contributors.
(d) Applications of OSM in different domains.
(e) Mining value-added knowledge and information from OSM.
(f) Limitations in the analysis of OSM data.
(g) Integration of OSM with commercial and non-commercial geodatasets.

We expect that this book will deliver significant scientific outcomes, which will further stimulate international research networking and collaboration. As outlined above, the inherent cross-disciplinary essence of OSM research combined with the emerging data quality, data mining, and patterns determination approaches to analysis of OSM means that contributing scholars for this book will be expected to have a diverse academic background not limited to geographic information science, cartography, computer science, statistics, and sociology. We feel that these trans- and inter-disciplinary contributions permit a deeper understanding of how the OSM project works and has become the phenomenal success that it is today. Last, but not least, the book will strive to bring OSM into the core of GIScience where the diverse worlds of new and classical geography and cartography will meet.

This book presents some cutting-edge developments and applications in the field of geography, spatial statistics, geographic information science, social science, and cartography. This collection of chapters is highly relevant for, but not limited to, the following potential audience and readership: researchers, postgraduates, and professionals.

The high response to our call for chapters shows that the intention of this book to be widely announced has been fulfilled. By the end of January 2014, a total of 34 chapter proposals were submitted and after an internal review by the editors, 30 authors of those originally 34 submitted chapter proposals were invited to submit a full chapter manuscript. After the final chapter submission deadline on 30 May 2014, a total of 29 manuscripts were submitted. Thereafter, each of the 29 chapter manuscripts was evaluated through a double-blind review process by at least two or three international experts in the respective field. For the review process, the standard Springer review guidelines were applied. Besides the innovative aspect of the research, the scientific quality of the research weighted heavily on the decision as to whether or not a manuscript was accepted or rejected. After two rounds of reviews conducted by international experts, the editors made the decision on whether or not the manuscript was fit for publication. In October 2014, 14 chapters were accepted and along with one introductory chapter and one conclusive chapter are now included in the present book.

5 Structure of the Book

The book covers several areas of OSM, each associated with a main theme of the book. The present volume has the following four sections: (1) Data management and quality, (2) Social context, (3) Network modeling and routing, and (4) Land management and Urban form. However, this structure should not be understood as fixed and definitive. Quite the contrary, the boundaries between these sections are partly fuzzy and overlap each other to some extent.

Section 1 on *Data Management and Quality* includes five chapters. In chapter "Assessment of Logical Consistency in OpenStreetMap Based on the Spatial Similarity Concept", Peyman Hashemi and Rahim Ali Abbaspour propose an approach for assessing the logical consistency in OSM based on the concept of spatial similarity in multi-representation considering three elements, i.e. directional relationships, topological relationships, and metric distance relationships. Jokar Arsanjani et al. in chapter "Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets" attempt to comparatively assess the accuracy of the contributed OSM land use features in four German metropolitan areas versus the pan-European GMESUA dataset as a reference. Their empirical findings suggest OSM to be alternative complementary source for extracting LU information with over half of the selected cities mapped by mappers. Moreover, the results identify which land types preserve high/moderate/low accuracy across cities for urban land use mapping. The findings strengthen the potential of collaboratively collected LU features for providing temporal LU maps as well as updating or enriching existing inventories. Chapter "Improving Volunteered Geographic Information Quality Using a Tag Recommender System: The Case of OpenStreetMap" by Arnaud Vandecasteele and Rodolphe Devillers proposes an approach for both improving the semantic quality and reducing the semantic heterogeneity of VGI datasets through implementing a tag recommender system, called OSMantic plugin, that automatically suggests relevant tags to contributors during the editing process. Their approach helps contributors find the most appropriate tags for a given object, hence reducing the overall dataset semantic heterogeneity. This plugin is developed for the Java OSM editor (JOSM) and different examples illustrate how this plugin can be used to improve the quality of VGI data. In chapter "Inferring the Scale of OpenStreetMap Features", Guillaume Touya and Andreas Reimer propose and compare two concurrent approaches at automatically assigning scale to OSM objects. Their first approach is based on a multi-criteria decision making model, with a rationalist approach for defining and parameterizing the respective criteria, yielding five broad Level of Detail classes. Their second approach attempts to identify a single metric from an analysis process, which is then used to interpolate a scale equivalence. Both approaches are combined and tested against well-known CORINE data, resulting in an improvement of the scale inference process. The chapter closes with a presentation of the most pressing open problems. In chapter "Data Retrieval for Small Spatial Regions in OpenStreetMap", Roland M. Olbricht investigates what design choices are required to be able to answer almost any geographic query whilst serving common use cases fast enough such that the services based on this database are fast on affordable and standard sized hardware. He evaluates the usage patterns from the main instance of Overpass API on overpass-api.de by considering more than 40 million requests from 2012 and 2013.

Section 2 deals with *Social Context* and comprises of three chapters. In chapter "The Impact of Society on Volunteered Geographic Information: The Case of OpenStreetMap", Afra Mashhadi et al. address whether the society and its characteristics such as the socio-economic factors have an impact on what part of the physical world is being digitally mapped so that we can understand where crowd-sourced map information can be relied upon. They measure the positional and thematic accuracy as well as the completeness of OSM data and quantify the role of society on the state of this digital production and finally quantify the effect of social engagement as a method of intervention for improving user participation. Georg Glasze and Chris Perkins frame a research agenda in chapter "Social and Political Dimensions of the OpenStreetMap Project: Towards a Critical Geographical Research Agenda" that draws upon critical cartography, but widens the scope of analysis to the assemblages of practices, actors, technologies, and norms at work: an agenda which is inspired by the critical GIS literature, to take the specific social contexts and effects of technologies into account, but which deploys a processual view of mapping. They recognize that a fundamental transition in mapping is taking place, and that OSM may well be of central importance in this process. In chapter "Spatial Collaboration Networks of OpenStreetMap", Klaus Stein et al. describe a new type of spatial collaboration network that can be extracted from OSM edit history data and show how to apply the measurement of interlocking responses known from research on non-spatial collaboration in wikis to collaboration in OSM. Finally, they discuss the advantages of their approach by demonstrating an analysis of collaboration on OSM sample data.

Section 3 deals with *Network modeling and routing* and includes three chapters. In chapter "Route Choice Analysis of Urban Cycling Behaviors Using OpenStreetMap: Evidence from a British Urban Environment", Godwin Yeboah and Seraphim Alvanides undertake a route choice analysis using the cycling-friendly version of OSM as the transportation network for analysis, alongside GPS tracks and travel diary data for 79 Utility Cyclists around Newcastle upon Tyne in North East England. They suggest that OSM can provide a robust transportation network for cycling research, in particular when combined with GPS track data, and conclude that network restrictions for both observed and shortest paths are significant, suggesting that route directness is an important factor to be considered for restricted and unrestricted networks. Chapter "The Next Generation of Navigational Services Using OpenStreetMap Data: The Integration of Augmented Reality and Graph Databases" by Pouria Amirian et al. describes the implementation of a navigational application as part of the eCampus project in Maynooth University. The application provides users with several navigation services with navigational instructions through standard textual and cartographic interfaces and also through augmented images showing way-finding objects. Jorge Gil presents the process of building a multi-modal urban network model using OSM data in chapter "Building a Multimodal Urban Network Model Using OpenStreetMap Data for the Analysis of Sustainable Accessibility". He develops various algorithmic procedures to produce the network model, supporting the reproducibility of the process and addressing the challenges of using OSM data for this purpose and addresses the great potential of OSM for urban analysis, thanks to the detail of its attributes and its open and universal coverage.

Land management and urban form is the focus of the final Sect. 4, comprising of three chapters. Chapter "Assessing OpenStreetMap as an Open Property Map" by Mohsen Kalantari and Veha La provides an assessment of OSM as a crowdsourcing system in collecting and recording land tenure information using a case study in Victoria, Australia. Their chapter studies the completeness of the public property records in OSM, and the location, shape, area and description of the existing records, and finally discusses the potential of OSM as an Open Property Map. Jacinto Estima and Marco Painho in chapter "Investigating the Potential of OpenStreetMap for Land Use/Land Cover Production: A Case Study for Continental Portugal" review the existing literature on using OSM data for land use/cover database production and move this research forwards by exploring the suitability of the OSM Points of Interest. They conclude that OSM can greatly contribute to mapping specific land types. In chapter "Using Crowd-Sourced Data to Quantify the Complex Urban Fabric—OpenStreetMap and the Urban–Rural Index”, Johannes Schlesinger presents an Urban–Rural Index (URI), which tackles the lack of any classification of the urban–rural continuum especially in regions of the world where accurate and up-to-date geodata is hardly available. His paper draws on the analysis of three study sites: Bamenda in Cameroon, Moshi in Tanzania, and Bangalore in India, and concludes that URI as a reproducible representation of the spatial complexity of the urban landscape and its surrounding areas has the potential to contribute to the understanding of urban development patterns.

In the final chapter (chapter “An Outlook for OpenStreetMap”) by Peter Mooney, the future research perspective of research on OpenStreetMap is reviewed. In the chapter, he structures his future vision of OpenStreetMap research by using the content of this volume and other OSM literature as a basis for future work.

Last but not least, the editors express their appreciation to all reviewers for their kind support and their critical and constructive comments for each chapter. Undoubtedly, their efforts have significantly enriched the quality of the entire volume. We profoundly appreciate the efforts of all authors who submitted a full chapter manuscript and chose this book as their desired publication outlet for their research. Furthermore, we would like to thank Prof. Muki Haklay for his invited comment. Finally, Jamal Jokar Arsanjani thanks the Alexander von Humboldt foundation and the Institute of Geography at Heidelberg University, Germany, for providing the foundation for this book. Peter Mooney would like to thank the Environmental Protection Agency Ireland for funding support. Finally, we acknowledge the Springer team as well as the series editors William Cartwright, Georg Gartner, Liqiu Meng, and Michael Peterson for their great assistance throughout the whole publication process. Certainly, without all these helping hands, this volume would have never been published. The book is partially sponsored by the COST Action IC1203.

References

Amelunxen C (2010) On the suitability of volunteered geographic information for the purpose of geocoding. In: Car A, Griesebner G, Strobl J (eds) Proceedings of the geoinformatics forum salzburg, na. Retrieved http://koenigstuhl.geog.uniheidelberg.de/publications/2010/Amelunxen/GI_Forum_Amelunxen.pdf

Bakillah M, Andrés Domínguez J, Zipf A, Liang SHL, Mostafavi MA (2013) Multi-agent evacuation simulation data model with social considerations for disaster management context. In: Zlatanova S, Peters R, Dilo A, Scholten H (eds) Intelligent systems for crisis management, SE-1. Springer, Berlin, Heidelberg, pp 3–16. doi:10.1007/978-3-642-33218-0_1

Bakillah M, Liang S, Mobasheri A, Jokar Arsanjani J, Zipf A (2014a) Fine-resolution population mapping using OpenStreetMap points-of-interest. Int J Geogr Inf Sci 28:1940–1963

Bakillah M, Lauer J, Liang S, Zipf A, Jokar Arsanjani J, Loos L, Mobasheri A (2014b) Exploiting big VGI to improve routing and navigation services. In: Big data techniques and technologies in geoinformatics, pp 177–192

Fan H, Zipf A, Fu Q, Neis P (2014) Quality assessment for building footprints data on OpenStreetMap. Int J Geogr Inf Sci 28:700–719. doi: 10.1080/13658816.2013.867495

Goetz M (2013) Towards generating highly detailed 3D CityGML models from OpenStreetMap. Int J Geogr Inf Sci 27:845–865

Goetz M, Zipf A (2012) Using crowdsourced indoor geodata for agent-based indoor evacuation simulations. ISPRS Int J Geo-Inf 1(2):186–208. doi:10.3390/ijgi1020186

Goodchild MF (2013) The quality of big (geo) data. Dialogues Hum Geogr 3:280–284

Graham M, Hale S, Stephens M (2012) Digital divide: the geography of internet access. Environ Plan A 44:1009–1010

Hagenauer J, Helbich M (2012) Mining Urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks. Int J Geogr Inf Sci 26:963–982

Helbich M, Amelunxen C, Neis P (2012) Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata. In: Jekel T, Car A, Strobl J, Griesebner G (eds) Geospatial crossroads @ GI_Forum 2012, Wichmann, Heidelberg, Salzburg, pp 24–33

Helbich M, Hagenauer J, Leitner M, Edwards R (2013) Exploration of unstructured narrative crime reports: an unsupervised neural network and point pattern analysis approach. Cartogr Geogr Inf Sci 40:326–336

Jokar Arsanjani J, Barron C, Bakillah M, Helbich M (2013a) Assessing the quality of OpenStreetMap contributors together with their contributions. In: 16th AGILE international conference on geographic information science, Leuven, Belgium, pp 14–17

Jokar Arsanjani J, Helbich M, Bakillah M, Hagenauer J, Zipf A (2013b) Toward mapping land-use patterns from volunteered geographic information. Int J Geogr Inf Sci 27: 2264–2278

Jokar Arsanjani J, Vaz E, Bakillah M, Mooney P (2014) Towards initiating OpenLandMap founded on citizens' science: the current status of land use features of OpenStreetMap in Europe. In: Huerta, Schade, Granll (eds) Proceedings of the AGILE'2014 international conference on geographic information science, Castellón, AGILE Digital Editions: Castellón, Spain, 3–6 June 2014

Jokar Arsanjani J, Helbich M, Bakillah M, Loos L (2015a) The emergence and evolution of OpenStreetMap: a cellular automata approach. Int J Digit Earth 8(1):74–88. doi:10.1080/17538947.2013.847125

Jokar Arsanjani J, Bakillah M, Arsanjani JJ (2015b) Understanding the potential relationship between the socio-economic variables and contributions to OpenStreetMap. Int J Digit Earth 0:1–16. doi:10.1080/17538947.2014.951081

Jokar Arsanjani J, Vaz E (2015c) An assessment of a collaborative mapping approach for exploring land use patterns for several European metropolises. Int J Appl Earth Obs Geoinf 35:329–337. doi:10.1016/j.jag.2014.09.009

Klonner C, Barron C, Neis P, Höfle B (2014) Updating digital elevation models via change detection and fusion of human and remote sensor data in Urban environments. Int J Digit Earth 1–19

Lauer J, Richter L, Ellersiek T, Zipf A (2014) TeleAgro+ - Analysis framework for agricultural telematics data, IWCTS '14, SIGSPATIAL'14, 4–7 Nov 2014, Dallas/Fort Worth, TX, USA

Mooney P, Corcoran P (2012) Characteristics of heavily edited objects in OpenStreetMap. Future Internet 4:285–305

Mooney P, Corcoran P (2013a) Analysis of interaction and co-editing patterns amongst OpenStreetMap contributors. Trans GIS

Mooney P, Corcoran P, Ciepluch B (2012) The potential for using volunteered geographic information in pervasive health computing applications. J Ambient Intell Humaniz Comput 1–15 (LA—English)

Mooney P, Rehrl K, Hochmair H (2013) Action and interaction in volunteered geographic information: a workshop review. J Locat Based Serv 7:291–311

Neis P (2014) Measuring the reliability of wheelchair user route planning based on volunteered geographic information. Trans GIS n/a–n/a, doi:10.1111/tgis.12087

Neis P, Zipf A (2012) Analyzing the contributor activity of a volunteered geographic information project—the case of OpenStreetMap. ISPRS Int J Geo-Inf 1:146–165

Neis P, Singler P, Zipf A (2010) Collaborative mapping and emergency routing for disaster logistics–case studies from the haiti earthquake and the UN Portal for Afrika. In: Proceedings of Geospatial Crossroads @ GI_Forum '10, Salzburg, Austria, 6–9 July 2010

Over M, Schilling A, Neubauer S, Zipf A (2010) Generating web-based 3D city models from OpenStreetMap: the current situation in Germany. Comput Environ Urban Syst 34:496–507

Roick O, Hagenauer J, Zipf A (2011) OSMatrix—grid-based analysis and visualization of OpenStreetMap. In: State of the Map EU 2011, Vienna, Austria

Rylov M, Reimer A (2014) A comprehensive multi-criteria model for high cartographic quality point-feature label placement cartographica. Int J Geogr Inform Geovis, doi:10.3138/carto.49.1.2137

Rylov M, Zipf A (2012) Solutions for limitations in label placement in OGC symbology encoding (SE) specification. In: Geoinformatik 2012, Heidelberg

Schilling A, Over M, Neubauer S, Neis P, Walenciak G, Zipf A (2009) Interoperable location based services for 3D cities on the Web using user generated content from OpenStreetMap. In: UDMS 2009. 27th Urban Data Management Symposium. Ljubljana , Slovenia

Schmitz S, Zipf A, Neis P (2008) New applications based on collaborative geodata—the case of routing. In: Proceedings of XXVIII INCA International Congress on Collaborative Mapping and Space Technology

Zielstra D, Zipf A (2010) Quantitative studies on the data quality of OpenStreetMap in Germany

Part I
Data Management and Quality

Assessment of Logical Consistency in OpenStreetMap Based on the Spatial Similarity Concept

Peyman Hashemi and Rahim Ali Abbaspour

Abstract The growth in the number of users and the volume of information in OpenStreetMap (OSM) indicate the success of this VGI-based project in attracting diverse sets of people from all over the world. A huge amount of information is generated daily by non-professional users and OSM faces the challenge of ensuring data quality. Spatial data quality comprises several basic elements; among them, logical consistency concerns the existence of logical contradictions within a dataset. It is one of the most important elements, but has not been studied much in VGI despite the key role in quality assurance. Because of the participatory nature of data collection and entry in OSM, the common consistency checking routines for spatial data should be revised. Since contributors have different views about objects, data integration in OSM may be considered as a form of multi-representation data combination. In this article, the concept of spatial similarity in multi-representation considering three elements, i.e. directional relationships, topological relationships, and metric distance relationships, is used to build a framework to determine the probable inconsistencies in OSM.

Keywords Volunteered geographic information (VGI) · Logical consistency · Spatial similarity · OSM · Topological relations

1 Introduction

The technological advancements in the geospatial domain such as developments in Web 2.0 mapping provide great opportunities to take advantage of a mass of volunteered users for geospatial data acquisition and enrichment. The well-known

P. Hashemi · R. Ali Abbaspour (✉)
Surveying and Geomatics Engineering Department, College of Engineering,
University of Tehran, Tehran, Iran
e-mail: abaspour@ut.ac.ir

P. Hashemi
e-mail: hashemipeyman@ut.ac.ir

J. Jokar Arsanjani et al. (eds.), *OpenStreetMap in GIScience*,
Lecture Notes in Geoinformation and Cartography,
DOI 10.1007/978-3-319-14280-7_2

projects such as OSM, Wikimapia, and Google Mapmaker, which are based on this idea, enable the users to play the role of geospatial data providers. The rising rate of users of such environments is an indication of the level of attraction. For example, the registered number of OSM members was 100,000 in 2009 while this reached over 1,500,000 by the beginning of 2014 (OSM statistics 2014).

Although VGI has proved a successful way to obtain detailed geographical information in a timely and low-cost manner, it suffers from some serious weaknesses (Goodchild and Li 2012). The Achilles' heel of VGI is the lack of metadata on spatial data quality parameters used for quality assurance mechanisms. Despite the professional surveyors, most of the users of VGI-based systems are unfamiliar with the standards on spatial data collection or they may feel no need to follow such rules. This leads to a serious problem when quality control procedures are employed to measure the quality level of spatial data. Thus, compatible, customized spatial data quality parameters and control procedures are required to insure the users to utilize those data by crossing the first step of collection of raw spatial data to spatial information production. The other reason to address the quality issues in VGI is the current movement from only data collection and enrichment towards volunteered geographic services (Sui et al. 2013).

There are several classifications of spatial data quality (Worboys and Duckham 2004; Morrison 1995; ISO 19157:2013), which are the results of efforts started in the 1980s in the geospatial information community. In almost all of them, five main elements of spatial accuracy, attribute accuracy, currency, logical consistency, and completeness are considered. They are more or less utilized in VGI studies to address the quality and uncertainty issues. Among the different elements of spatial data quality, logical consistency is studied less than the others in VGI research. Due to the numerous topological errors and loss of standards, which should be considered by users during data entry, different types of inconsistencies occur in VGI. Although there are mechanisms to discover some inconsistencies such as open polygons in some VGI projects like OSM, other basic errors such as overshoots and undershoots, which are of high frequency, remain unsolved. Therefore, it is necessary that the logical consistency, especially with a focus on topological aspects, be assessed in order to fulfill the spatial quality assurance in VGI.

Literature on the quality study in VGI, especially in OSM, may be generally divided into three categories. In the first category, the data in OSM is compared with reference data, which have higher precision and are generated by a mapping agency (Haklay 2010; Girres and Touya 2010; Zielstra and Zipf 2010; Neis et al. 2011). Comparisons of volunteered spatial data with reference data constitute a considerable portion of the research in this field. There are several reports of evaluations of this type which have been carried out in various countries such as the U.K. (Haklay 2010), France (Girres and Touya 2010), and Germany (Zielstra and Zipf 2010; Neis et al. 2011). Although this approach looks rational, using it in most cases in VGI, where there are no reference data, is impossible. Moreover, the costs and license limitations restrict access to high quality, commercial datasets. The research in the second category concentrates on user activities (Jokar Arsanjani et al. 2013), and the third category is comprised of research on the history of the

```xml
<node id="27128507" lat="50.0146754" lon="8.2429516" version="1" timestamp="2007-04-05T18:11:17Z"
  changeset="6911" uid="3609" user="seb"> <tag k="created_by" v="JOSM"/> </node>
<node id="27128507" lat="50.014677" lon="8.2429489" version="2" timestamp="2007-10-16T12:38:56Z"
  changeset="203178" uid="16643" user="Joh"/> </node>
<node id="27128507" lat="50.0144034" lon="8.2431315" version="7" timestamp="2013-03-16T23:41:50Z"
   changeset="15390280" uid="440308" user="spezialist"/> </node>
```


Fig. 1 A sample of the OSM history file

data (Keßler and de Groot 2013; Keßler et al. 2011). Data recorded as the history files in OSM data can play an alternative role as reference data. All OSM data such as nodes, ways, and relations as well as their previous versions are stored in the full history dump (OSM, full history dump 2014). These files contain information such as node position (latitude and longitude), object versions, modified times (timestamp), user identification (user name and id), and tags information (key and value). For instance, the submitted information for a node in the OSM history files, which is actually a gas station, is shown in Fig. 1. This file is in the XML format. This comprehensive history file allows researchers to conduct more statistical analysis on spatial aspects or a contributor's information. Thus, intrinsic quality indicators can be applied instead of reference datasets (Barron et al. 2013).

This paper addresses the logical, topological inconsistencies in VGI with a focus on the popular OSM project. A similarity-based framework is used to deal with this issue. The remainder of this paper is organized as follows: First, there is an introduction to logical consistency in general and then the focus is on the meaning of this concept in OSM. Afterwards, the proposed framework is introduced and its main components are described in detail. The last part of this section is the explanation of the proposed methodology and its evaluation. The final section summarizes the work.

2 Logical Consistency for OSM

The concept of logical consistency was initially introduced for database integration purposes (Kainz 1995). Logical constraints defined for a database are rules that adapt the data to the selected structure and provide the opportunity to optimize the storage/retrieval speed. This concept is then used by the geospatial community to address the different sorts of inconsistencies arising during data entry and analysis in GIS. Logical consistency is highly correlated with positional errors. Because there are numerous sorts of positional errors in VGI-based projects, the logical inconsistencies are of high frequency in these projects, which in turn could have a side effect on the analysis and usage of the information.

Based on ISO 19157:2013 standard, logical consistency can be defined as "degree of adherence to logical rules of data structure, attribution, and relationships". Four main sub-elements considered for logical consistency in this standard are conceptual consistency, domain consistency, format consistency, and topological consistency. Conceptual consistency monitors the adherence to the rules of the conceptual schema. Domain consistency checks the value ranges, which should be in certain value