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

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

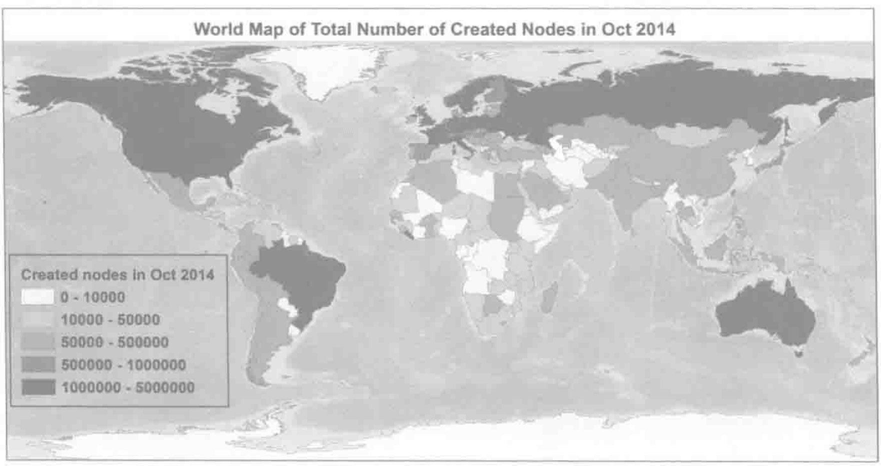

Fig. 5 A world map of the total created nodes in October 2014

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.

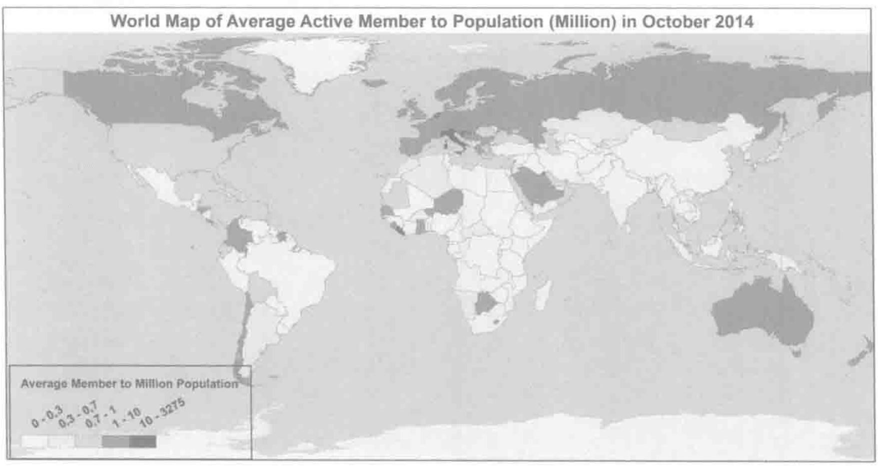

Fig. 6 A world map of average number of active members to population (million) in October 2014