Volunteered Geographic Information and the Future of Geospatial Data

Cláudio Elízio Calazans Campelo
Federal University of Campina Grande, Brazil

Michela Bertolotto
University College Dublin, Ireland

Padraig Corcoran
Cardiff University, UK

A volume in the Advances in Geospatial Technologies (AGT) Book Series

IGI GLOBAL
DISSEMINATOR OF KNOWLEDGE

www.igi-global.com

Published in the United States of America by
IGI Global
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

CIP Data Pending
ISBN: 978-1-5225-2446-5
eISBN: 978-1-5225-2447-2

This book is published in the IGI Global book series Advances in Geospatial Technologies (AGT) (ISSN: 2327-5715; eISSN: 2327-5723)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 2
Quality Evaluation of Volunteered Geographic Information:
The Case of OpenStreetMap

Hongyu Zhang
Western University, Canada

Jacek Malczewski
Western University, Canada

ABSTRACT

A large amount of crowd-sourced geospatial data have been created in recent years due to the interactivity of Web 2.0 and the availability of Global Positioning System (GPS). This geo-information is typically referred to as volunteered geographic information (VGI). OpenStreetMap (OSM) is a popular VGI platform that allows users to create or edit maps using GPS-enabled devices or aerial imageries. The issue of quality of geo-information generated by OSM has become a trending research topic because of the large size of the dataset and the inapplicability of Linus' Law in a geospatial context. This chapter systematically reviews the quality evaluation process of OSM, and demonstrates a case study of London, Canada for the assessment of completeness, positional accuracy and attribute accuracy. The findings of the quality evaluation can potentially serve as a guide of cartographic product selection and provide a better understanding of the development of OSM quality over geographic space and time.

INTRODUCTION

Although a large amount of geospatial data and wide range of applications have made GIS very popular, the users are often unaware of the data quality. New elements were added to the discussion of geospatial data quality in the 21st century through the development of Web 2.0 and the availability of Global Positioning System (GPS). The interactivity of the new web technology helped create a large amount of user-generated content (UGC). UGC with location information is referred to as user-generated geospatial

DOI: 10.4018/978-1-5225-2446-5.ch002

content (Coleman, Georgiadou, & Labonte, 2009), crowd-sourced geodata (Barron, Neis, & Zipf, 2014) or volunteered geographic information (VGI) (Goodchild, 2007). More specifically, using location-based services (LBS), GPS-enabled devices and/or satellite images, VGI users actively upload and share data following an opt-in provision, and VGI can be direct or indirect depending on whether users have local knowledge (Haklay, 2013). The activities of contributing VGI have been termed in different ways as well, including collaborative mapping (Jokar Arsanjani & Vaz, 2015), participatory GIS (Elwood, 2006) and public participation GIS (PPGIS) (Lin, 2013).

Researchers are interested in VGI because of its values. The conventional apprehension about commercial or governmental cartographical products is authoritative, comprehensive and accurate. However, Coleman (2013) and Dobson (2013) concluded that these databases are often out-of-date, incomplete, of inconsistent quality, and costly to maintain. Therefore, VGI is studied as a crowd-sourced alternative to "authoritative" datasets. OpenStreetMap (OSM) is one of the VGI applications that allow users to create and edit maps using satellite images. As of July 2016, the total number of registered users on OSM has passed 2.8 million, creating more than 3.4 billion nodes (data points) accumulatively ("OSMstats - Statistics of the free wiki world map," 2016). This chapter systematically summarizes the quality evaluation process of OSM through literature review and a case study in London, Canada, with focuses on the comparisons of different assessment methods and findings.

BACKGROUND

The term volunteered geographic information (VGI) was suggested by Goodchild (2007) to represent geospatial data contributed by individuals voluntarily. Since VGI is often the most cost-effective solution, the crowd-sourced geodata have been applied in many fields such as participatory planning and spatial decision making. Moreover, VGI is the only source of geodata in some regions because of security or financial concerns. The area of humanitarian relief and crisis management is the most prominent application of VGI. Ushahidi and the Humanitarian OpenStreetMap Team (HOT) are two platforms that have had strong presence on disaster management since 2008 and 2009 respectively. Table 1 compares some VGI applications with OSM. Although OSM is not the project with the longest history, it is the

Table 1. Comparison of volunteered geographic information (VGI) applications

Attributes	OpenStreet Map	Wikimapia	Waze	Moovit	GasBuddy
Founding year	2004	2006	2008	2012	2000
Specialization	Mapping	Mapping	Navigation	Public transit	Fuel prices
Number of users or registered members (in million)	2.8 (in 2016)	1.9 (in 2013)	50 (in 2013)	20 (in 2014)	35 (n.d.)
Coverage in 2016	World	World	World	600+ cities	United States and Canada
License	ODbL	CC BY-SA	Proprietary	Proprietary	Proprietary
Data downloadable	Yes	Yes	No	No	No

Note. ODbL, Open Data Commons Open Database License; CC BY-SA, Creative Commons license Attribution-ShareAlike; data for OpenStreetMap from "OSMstats - Statistics of the free wiki world map" (2016), for Wikimapia from Neis & Zielstra (2014), for Waze from CBC News (2013), for Moovit from "Moovit Company Overview" (2014), and for GasBuddy from "Advertise with us - Gasbuddy Gas Prices" (n.d.).

oldest mapping project in which the geo-information can be applied in more than one field. The number of “registered members” of OSM is relatively small comparing to other specialized applications, but the number of “users” could be a bloated figure and does not represent “active contributors”. Jokar Arsanjani and Bakillah (2014), Mooney and Corcoran (2012a) and Yang, Fan, and Jing (2016) provide more insights into the OSM contribution patterns and user behavior. Like Wikimapia and Waze, OSM has a worldwide coverage. The difference is that OSM allows users to freely alter and redistribute its data, which is accessible through multiple servers in different formats. In contrast, Wikimapia only offers its data by a web application programming interface (API) (Neis & Zielstra, 2014), and Waze does not release data from its platform. Therefore, OSM was chosen to be the focus of this chapter. The following subsections starts with the discussion of quality concerns in VGI, introduces OSM in details and ends with a list of the spatial data quality metrics.

Quality Issues of Volunteered Geographic Information (VGI)

Community-based systems, like the review system on Amazon or Airbnb, could be useful to evaluate the relative and latent value of VGI (Feick & Roche, 2013). Data quality assessment is a more explicit way of determining the value of VGI. Quality issues of VGI are typically centered around inconsistency in terms of coverage and accuracy. For instance, remote areas are usually under-mapped (Coleman, 2013). If volunteers are unfamiliar with the remote areas they map, accuracy might be sacrificed because of volunteers’ deficiency of local knowledge (Dobson, 2013). In addition to geometrical objects, VGI’s metadata is also incomprehensive and inaccurate (Hashemi & Ali Abbaspour, 2015), which creates difficulties for researchers to verify the semantic accuracy of VGI. The demographic composition of VGI contributors has some patterns as well; for example, Haklay (2013) suggested that VGI contributors were mainly male with high income, and Neis and Zipf (2012) found that 72% of OSM users was located in Europe. It is still unclear how this composition influenced VGI quality. Although the International Organization for Standardization (ISO) has published quality principles for geographic information (ISO, 2002), a new quality assurance schema specifically tailored for VGI is needed because of the limitations mentioned above (Van Exel, Dias, & Fruijtier, 2010).

The simplified expression of Linus’ Law – “Given enough eyeballs, all bugs are shallow” (Raymond 2001, p. 13) – is often quoted as an underlying theory for discussing the issues of data quality (e.g., Haklay et al. 2010; Miller and Goodchild 2015; Goodchild and Li 2012; Goodchild 2013). However, Linus’ Law may not work well in a spatial context (Elwood, Goodchild, & Sui, 2013), and this quotation often misleads readers to conclude that most quality issues will be solved if there are enough testers. The full expression of Linus’ Law is that “Given a large enough betatester and codeveloper base, almost every problem will be characterized quickly and the fix obvious to someone” (Raymond 2001, p. 13). This expression specifies that the “eyeballs” must include those from co-developers, who are professionally trained to debug the Linux operating system in the context of the Raymond’s article. However, VGI contributors are mainly citizen scientists but not professional cartographers. Moreover, the software “bugs” can be identified during the process of using the software. However, errors on maps cannot be recognized or avoided if map scale is too small, contributors do not have local knowledge, or accuracy is sufficient for certain map applications (i.e., navigation requires less accuracy than road constructions). Furthermore, the contribution pattern of VGI users signifies the necessity of spatial redundancy (Dobson, 2013). For example, 38% of registered OSM members edited at least once, and only 5% of all actively contributed to the project (Neis & Zipf, 2012). Spatial heterogeneity also prevents the existence of consistent spatial errors across the globe that may be corrected all at once. Thus, Linus' Law does not apply to VGI, which means a large number of volunteers may not be enough to ensure the quality of VGI.

OpenStreetMap (OSM)

OSM is a crowdsourced online mapping platform, which aims to provide free and editable digital mapping products under a new copyright license (Haklay & Weber, 2008). Since its initiation in 2004, OSM has been applied in routing and navigation, cartography improvement, Location Based Services (LBS), and 3D city models (Jokar Arsanjani, Zipf, Mooney, & Helbich, 2015). In 2014, high densities of OSM nodes were found in North America, Europe, Russia, Australia and Brazil, while Africa and Greenland were least mapped (Jokar Arsanjani, Zipf, et al., 2015). According to Jokar Arsanjani, Zipf, et al. (2015), Mooney & Corcoran (2014), Neis & Zielstra (2014), Stein, Kremer, & Schlieder (2015) and Vandecasteele & Devillers (2015), OSM can be described by the following key features:

- **Near Real-Time Updates:** Unlike Google Map Maker, which has a review system for submitted edits, OSM publishes modifications just "a few minutes" after contributors save changes;
- **Data Import From Multiple Sources:** OSM supports data generated from Global Positioning System (GPS), smartphones, and other mapping hardware. In the early years of the project, GPS-enabled devices were the most popular data generators. This situation was changed because Yahoo! (from 2007 to 2011) and Microsoft Bing (since 2010) agreed to provide their aerial imageries for OSM enthusiasts to trace data. Some countries such as the United States and Canada also had volunteers to import authoritative datasets into OSM;
- **Data Export in Multiple Formats:** OSM data can be downloaded at different scale (e.g., continental, regional or metro) in different formats (e.g., OSM Extensible Markup Language (XML), Protocol Buffer Binary Format (PBF) or shape file) from several servers (e.g., Planet OSM, Geofabrik or Mapzen);
- **Different Flavours of Editors:** The web-based iD editor has a simple user interface for beginners to immerse into geodata contributions. Besides, Potlatch or JOSM (Java OpenStreetMap Editor) are favoured by advanced mappers. Other editors are available across operating systems and platforms as well;
- **Full Edit History:** OSM keeps all historical edits in its full history dump site ("Index of /planet/full-history," 2016), but only the latest object versions in other forms of extracts;
- **Three Different Object Types:** A "node" represents a point, while a "way" consists of lines or polygons (closed line features). A "relation" connects related nodes and ways with each other;
- **Tags as Metadata:** Attributes of objects are expressed as "key = value" pairs;
- **Spatial Heterogeneity:** Patterns of contributions differ from one place to another;
- **Manifold Collaboration Channels:** The official OSM wiki provides the knowledge base of the project. Other communication methods include Internet relay chats (IRCs) ("IRC - OpenStreetMap Wiki," 2015) and mailing lists ("Mailing lists - OpenStreetMap Wiki," 2016). Community events such as "mapping parties" are organized both online and offline, with the yearly "State of the Map" conference attracting most attendees.

Spatial Data Quality

Spatial data quality can be evaluated internally or externally (Jokar Arsanjani, Mooney, Zipf, & Schauss, 2015). While external quality assesses the fitness of data for a particular purpose, internal quality describes how well data meet specifications, such as:

- **Completeness (C):** Measures the comprehensiveness of a dataset. This criterion not only reports how much data is missing, but also the amount of data that should be excluded;
- **Positional Accuracy (PA):** Measures the relative and absolute accuracy of coordinate values;
- **Attribute Accuracy (AA):** Measures the correctness of attributes associated with geometrical shapes, which is also known as thematic accuracy (ISO, 2002);
- **Logical Consistency (LC):** Measures the internal consistency of a dataset, such as topological correctness and relations of objects;
- **Semantic Accuracy (SA):** Measures whether data objects and their meanings are interpreted correctly;
- **Temporal Quality (TQ):** Measures the validity of changes and the rate of updates in a dataset;
- **Lineage (L):** Measures the history of a dataset from collection to evolution (Van Oort, 2006).

The focus of this chapter is the internal quality of VGI data. However, it has been recognized that the above criteria only assess absolute data quality, while the actual quality is relative to its fitness-of-use (Feick & Roche, 2013; Van Oort, 2006).

A REVIEW OF OPENSTREETMAP QUALITY ASSESSMENT

A systematic survey of literature (as of July 2016) found 60 articles relevant to quality evaluation of OSM (see Appendix). Four databases were used in this process including Web of Science, Scopus, Engineering Village (Geobase) and Proquest (dissertations & theses). 334 articles were found initially using keywords "OpenStreetMap AND (quality OR accuracy)" with the option of anywhere except full text, and the number of relevant articles went down to 202 after removing duplicates. A full-text review of the 202 articles identified 39 articles listed in the Appendix. In addition, 21 relevant articles were found based on an examination of the 39 articles' reference sections. Only studies written in English were retained. It is worth to mention that some excluded articles are not totally irrelevant, but they focus more on method assessment instead of quality of specific areas (Basiri et al., 2016; Brovelli, Minghini, Molinari, & Mooney, 2016; Fan, Yang, Zipf, & Rousell, 2015; Graser, Straub, & Dragaschnig, 2014; Gröchenig, Brunauer, & Rehrl, 2014; Jokar Arsanjani, Mooney, Helbich, & Zipf, 2015; Zhang & Ai, 2015). In Table 10 of the Appendix, time represents the actual time the OSM data was downloaded, which is more accurate than the year of publication. Only years were recorded because of various time precision. Data were retrieved from 2007 to 2014, indicating the discussion of OSM quality assessment started around 2007 and continued as a trending topic until recent times. A limited number of studies were implemented using national data, signifying current exploration stage of OSM quality analysis. Most studies had European regions as their study areas, which was not surprising considering the massive number of European OSM users. Furthermore, most studies used a reference dataset to evaluate the extrinsic quality of OSM data, which include a mix of governmental and commercial databases. For articles that do not have a reference dataset, some constructed frameworks, some analyzed user behavior or data trust, and the rest studied intrinsic quality using data history.

The frequency of examined data quality criteria is shown in Figure 1. Data completeness dominates the quality analysis of OSM, with positional accuracy and attribute accuracy the second and the third most popular criterion. The common evaluation methods of all criteria are explained in the following paragraphs.

Generally, there are two types of methods to measure data completeness: unit-based and object-based (Table 2). The concept behind unit-based methods is to compare total length, area, or number of objects in OSM with those in a reference dataset. Many studies have used this method because of its easiness of implementation. Hochmair et al. (2015) specially considered street network density and visually compared bike lanes with Google street view to avoid potential mistakes. On the other hand, (automated) feature matching is involved in object-based methods using attributes or geometric properties. For example, street segments have orientation and length, and building footprints can be matched by their centroids or overlap ratio between OSM data and a reference. It is worth to mention that the completeness of land use may be calculated without a reference, since a 100% result means everywhere is covered by a land use feature (Jokar Arsanjani, Mooney, Zipf, et al., 2015).

The methods of measuring positional accuracy are categorized by data types (Table 3). A common method for points of interest is Euclidean distance, while buffer analysis is popular for line segments. A buffer of width “x” is created around a road segment from an authoritative dataset, and the percentage of the corresponding OSM road segment that falls within the buffer is calculated (Goodchild & Hunter, 1997). The buffer size differs from one study to another, indicating that there is no theory behind this method, and empirical analysis is the key to determine the buffer size. In terms of polygon features, centroids, corner points and surface are considered for distance measurements.

The methods of measuring attribute accuracy have four types of usages (Table 4). First, presence of OSM tags (e.g., oneway flags of street segments) can be looked up through examining each geometric object. Second, similarities of strings can be calculated by different algorithms. For example, the Lev-

Figure 1. Summary statistics of examined data quality criteria in Appendix

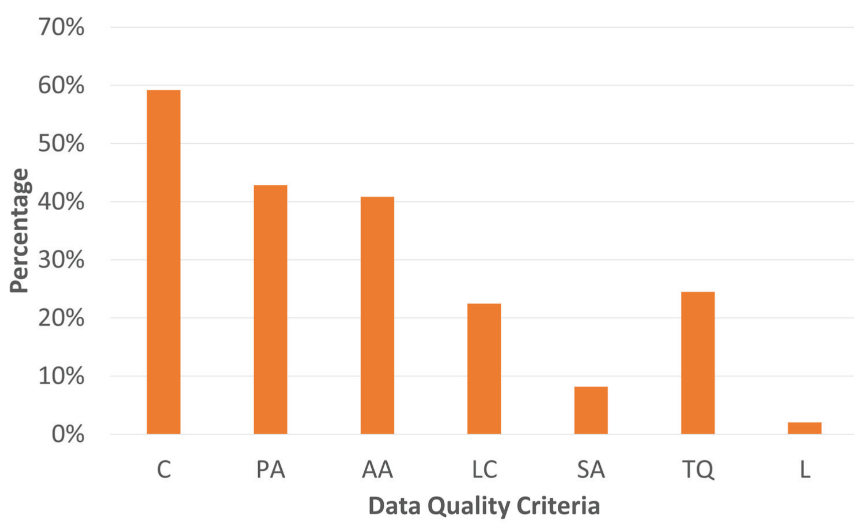


Table 2. Methods of measuring completeness

Types	Criteria	Examples
Unit-based	Number of objects (e.g., attributes, POIs or buildings)	Barron, Neis, & Zipf (2014), Fan, Zipf, Fu, & Neis (2014), Girres & Touya (2010), Haklay (2010), Hecht, Kunze, & Hahmann (2013), Hochmair, Zielstra, & Neis (2015), Jackson et al., (2013), Jokar Arsanjani, Barron, Bakillah, & Helbich (2013), Jokar Arsanjani, Mooney, Zipf, & Schauss (2015), Jokar Arsanjani & Vaz (2015), Mashhadi, Quattrone, & Capra (2015), Neis, Zielstra, & Zipf (2011), Zielstra & Zipf (2010)
	Total length or area	
	Density	Hochmair et al. (2015)
	Visual comparison	
Object-based	Centroids	Hecht et al. (2013)
	Overlap ratio	
	Attribute match (e.g., name)	Jackson et al. (2013), Kalantari & La (2015), Koukoletsos, Haklay, & Ellul (2012), Ludwig, Voss, & Krause-traudes (2011)
	Geometric match (e.g., distance, orientation, length)	

Table 3. Methods of measuring positional accuracy

Data Types	Methods	Examples
Point	Euclidean distance	Girres & Touya (2010)
		Amelunxen (2010)
		Jackson et al. (2013)
Line	Compare actual road conjunction with previous locations	Barron, Neis, & Zipf (2014)
	Hausdorff distance	Girres & Touya (2010)
	Average distance (McMaster, 1986)	
	Buffer analysis (Goodchild & Hunter, 1997; Hunter, 1999)	Haklay, (2010), Jokar Arsanjani, Barron, Bakillah, & Helbich (2013), Ludwig, Voss, & Krause-traudes (2011)
	Bidimentional regression (Friedman & Kohler, 2003; Tobler, 1994)	Helbich, Amelunxen, & Neis (2012)
	G*-statistics (Getis & Ord, 1992)	
Polygon	Surface distance (Vauglin, 1997)	Girres & Touya (2010)
	Average distance of corresponding (corner) points	Fan, Zipf, Fu, & Neis (2014)
	Distance between centroids	Kalantari & La (2015)

enshtein distance is the number of deletions, insertions, or reversals required to transform one string to another. The algorithm was originally developed to tackle the issue of binary information transmission (Levenshtein, 1966). The larger the Levenshtein distance, the greater the differences between strings. Third, numbers can be subtracted, and the absolute values of the results can reflect the differences between them. Finally, thematic accuracy (e.g., for land use accuracy assessment) can be measured by confusion matrix and kappa index.

A framework was constructed exclusively for logical consistency (Hashemi & Ali Abbaspour, 2015). Spatial scenes – sets of spatial objects with spatial relations – are compared in this framework. Topology, distance and direction are some examples of useful spatial relations (Hashemi & Ali Abbaspour, 2015). Here, topology is "the study of qualitative properties that are invariant under distortion of geometric space" (e.g., the London underground map) (Jiang, 2013, p. 128). For instance, two articles from Appendix studied logical consistency of street networks considering topological errors (e.g., connectivity of roads and structure of network), turn restrictions and inter-theme consistency (Girres & Touya, 2010; Neis et al., 2011). Another two articles examined logical consistency of polygons, both using shape similarity ratio in additional to other methods such as turning function distance, number of vertices, mean vertex spacing distance, and feature areas (Fan et al., 2014; Kalantari & La, 2015). Although OSM has a dedicated webpage to record known data errors ("Quality assurance - OpenStreetMap Wiki," 2016), Girres & Touya (2010) mentioned that integrity constraints are not enforced to ensure logical consistency in OSM.

Methods of other data quality criteria are summarized below. Only four out of the 60 articles analyzed semantic accuracy, and two of them compared attributes for the assessment (Girres & Touya, 2010; Jokar Arsanjani, Barron, et al., 2013). Fan et al. (2014) did something special to identify the n:m relations of building footprints between OSM data and a reference dataset. Temporal quality was generally evaluated as a spatial-temporal analysis with the rate and accuracy of changes over time. Level of Details (LOD) assessment can be divided into five schemas including conceptual schema, geometric resolution, semantic resolution, geometric precision and granularity (the size of the minimal features) (Touya & Reimer, 2015). Finally, a number of collected studies analyzed relations between user behaviors or data trust to user information and/or edit history.

Mixed results were found across different locations, times, data types and criteria. Some urban areas with high population density had similar or even better quality than some reference datasets. However, rural areas received less attentions and had scarce coverage. Overall, the findings of collected articles

Table 4. Methods of measuring attribute accuracy

Usages	Criteria	Examples
Measures attribute completeness	Tag presence	Girres & Touya, (2010), Ludwig et al. (2011)
Compares strings (text)	Levenshtein distance (Levenshtein, 1966)	Girres & Touya (2010)
	Similarity ratio (calculated by difflib in Python)	Kalantari & La (2015)
Compares numbers	Difference in speed limits	Ludwig et al. (2011)
Measures thematic accuracy	Classification accuracy by confusion matrix	Estima & Painho (2013), Jokar Arsanjani, Helbich, Bakillah, Hagenauer, & Zipf (2013), Jokar Arsanjani, Mooney, et al. (2015), Jokar Arsanjani & Vaz (2015)
	Kappa index	

follow the two classical geographic theories: Tobler's (1970) first law of geography – near things are more related than others – and the second law of geography – geographic phenomena vary across the globe (spatial heterogeneity) (Goodchild, 2009).

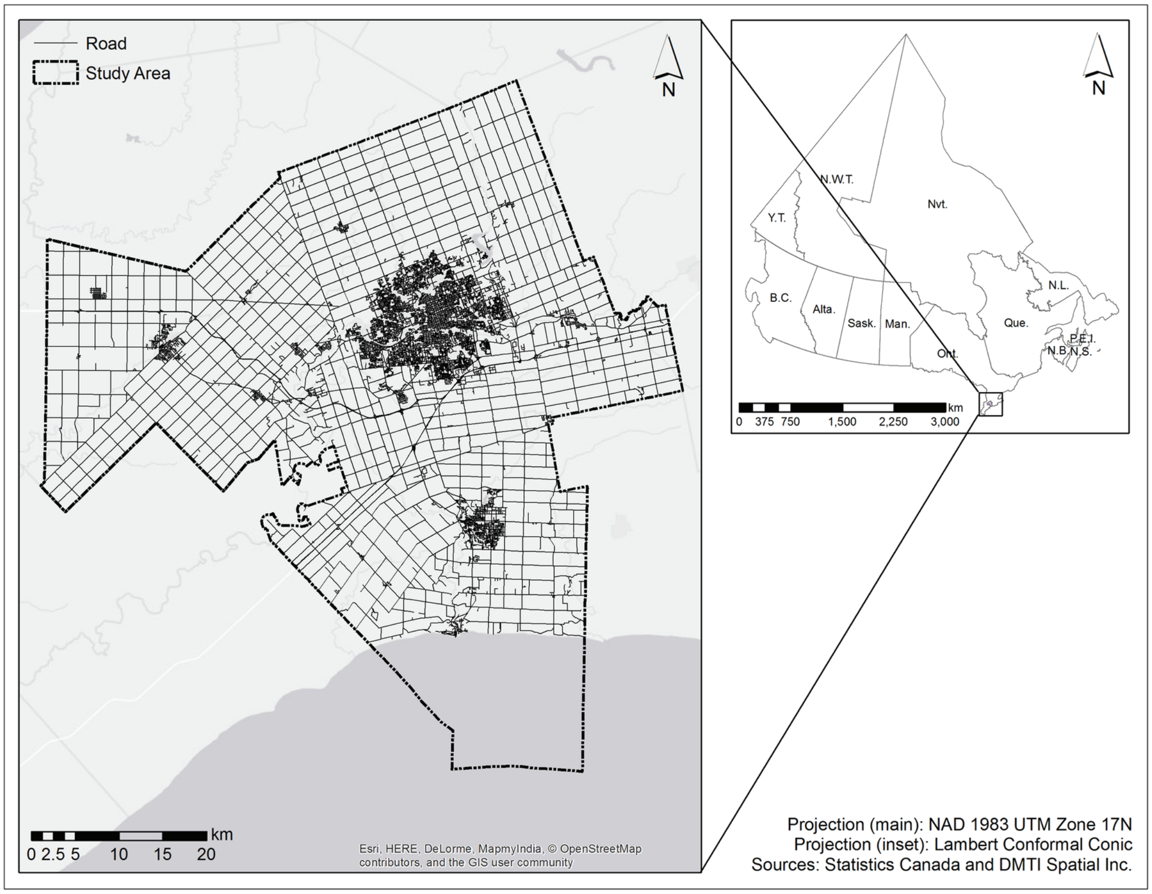
CASE STUDY

According to the Appendix, only a small number of articles evaluated quality of Canadian OSM data (e.g., Meier, 2015; Tenney, 2014). Although Tenney (2014) performed a national study, the results were still preliminary. Thus, there is a need to further evaluate the Canadian OSM quality. The study area here is the Census Metropolitan Area (CMA) of London, Ontario, Canada (Figure 2). London is the eleventh largest CMA in Canada with more than 474,000 inhabitants, including two cities (London and St. Thomas), two municipalities (Thames Centre and Central Elgin) and four townships (Strathroy-Caradoc, Middlesex Centre, Southwold and Adelaide-Metcalfe) (Statistics Canada, 2012). The rate of economic growth in the region was moderate in recent years because of an improved manufacturing sector and a stronger housing market. Two datasets, the source and the reference data, are required for this evaluation of OSM accuracy. The source data are the 2016 OSM metro extracts of London, Ontario from Mapzen[1] in the imposm format[2]. The reference data are the 2015 DMTI road networks from Scholars Geoportal[3], which has a positional accuracy ranging from 0.6 (urban) to 30 m (rural) (DMTI Spatial Inc., 2015). It is therefore hypothesized that urban roads have higher positional accuracy in OSM as well. The 2015 National Road Network (NRN) data collected by Natural Resources Canada is not chosen as the reference dataset because a commercial dataset is preferred when available (Haklay, 2010). The positional accuracy of the NRN data is not specified either (only indicated "in meters") (Natural Resources Canada, 2015).

Methods

The OSM quality, specifically completeness, positional accuracy and attribute accuracy, was assessed using the following techniques and ArcGIS tools (Figure 3 and 4). The attributes were first processed and matched based on Table 5. Evaluation results were classified according to the new road ranks in Table 6. Geometric feature matching was also performed before evaluating the positional and attribute accuracy. The unmatched road segments were identified using the "Detect feature changes" tool in ArcGIS with a search distance of 30 m (the maximum positional offset of the DMTI data) and removed afterwards (Figure 4). The length and density of roads were calculated to analyze the data completeness. This unit-based method was chosen because it is easy to implement and has been used in many previous studies (Table 2). Next, the buffer analysis was used to assess the positional accuracy (Figure 4). This method was validated in the first OSM quality assessment (Haklay, 2010) and other studies (Table 2). Using a self-developed python script and the arcpy library, buffers with widths of 1 to 10 m were created around the DMTI street networks, and the matched OSM road segments that fell within the buffers were clipped for calculating their proportions to the total OSM road length (Figure 5). Finally, the attribute accuracy was evaluated by tag presence, number difference and Levenshtein distance. Tag presence measured whether an OSM road attribute was present if a DMTI road attribute was provided (Figure 4). The absolute difference between two numeric fields were calculated as follows: $d = \left|x - y\right|$. Levenshtein distance (see the review section for definition) of two text fields was computed using a dynamic programming python script (Levenshtein, 1966).

Figure 2. Study area



Results and Discussion

Completeness

Figure 6 shows the road lengths by ranks. Many of the ranks have close lengths except rank 0, 5 and 6. Visual examination confirmed that most unclassified (rank 0) road segments of OSM are local roads (rank 5) in suburban areas. Thus, the length difference of rank 5 is actually minimal if the length of rank 0 is added. The difference of rank 6 is large enough to influence the total road lengths because of the large number of footways in the OSM data. This is also the case of the United States (as of 2012) (Zielstra, Hochmair, & Neis, 2013) and Germany (as of 2011) (Neis et al., 2011). If rank 6 is excluded, the difference is significantly reduced. However, OSM has a longer total length than DMTI with or without rank 6, which is different from previous studies since the total length of OSM motorways was still shorter than reference datasets (Neis et al., 2011; Zielstra et al., 2013). The better data completeness potentially benefits from data imports and the increased number of contributors over the years.

The road density of the two datasets is displayed in Figure 7. In general, urban areas especially the City of London and the City of St. Thomas have higher road density, which potentially helps to generate shorter and better routes in navigation applications (Mondzech & Sester, 2011). The location of dense

Figure 3. An overview of the methodology

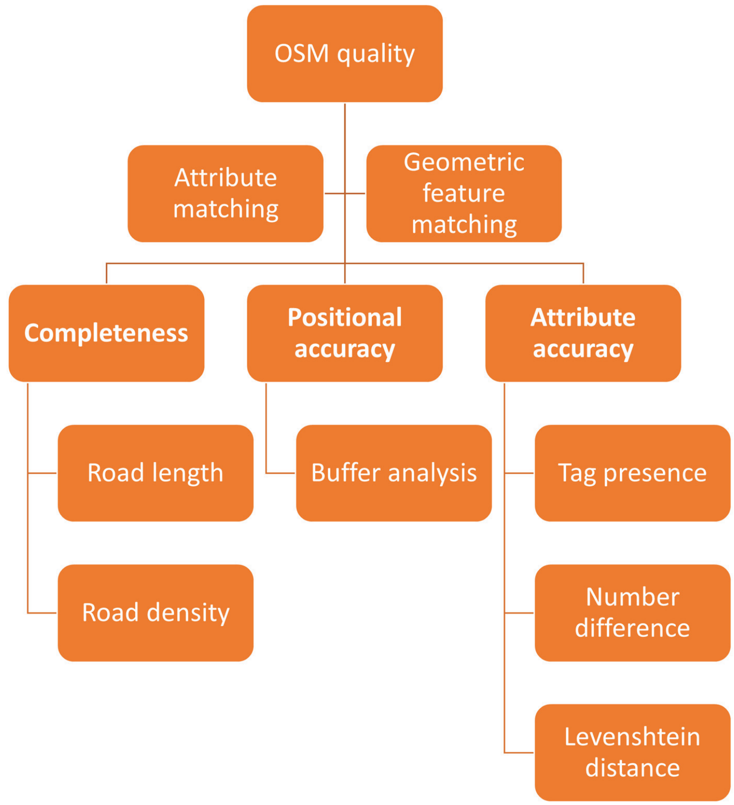
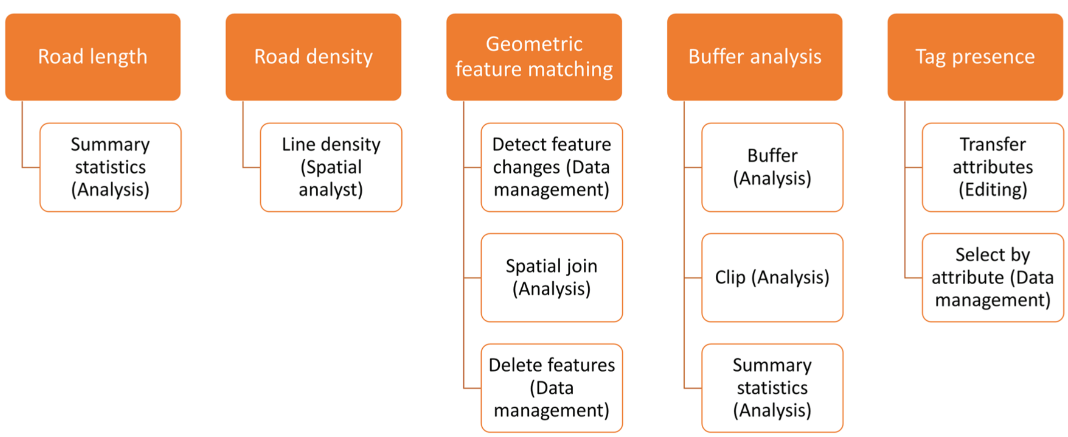


Figure 4. ArcGIS tools used for the evaluation



areas verifies that areas with denser population tend to have higher contributions (Jokar Arsanjani & Bakillah, 2014). The maximum density of DMTI is about 12 m/km^2, which is significantly less than the 30 m/km^2 of OSM. The difference is reflected in urban areas where the OSM map has a much darker color. The significant disparity of rank 6 should have great influence on the road density as well.

Figure 5. Example of the buffer analysis
(adapted from Goodchild & Hunter, 1997)

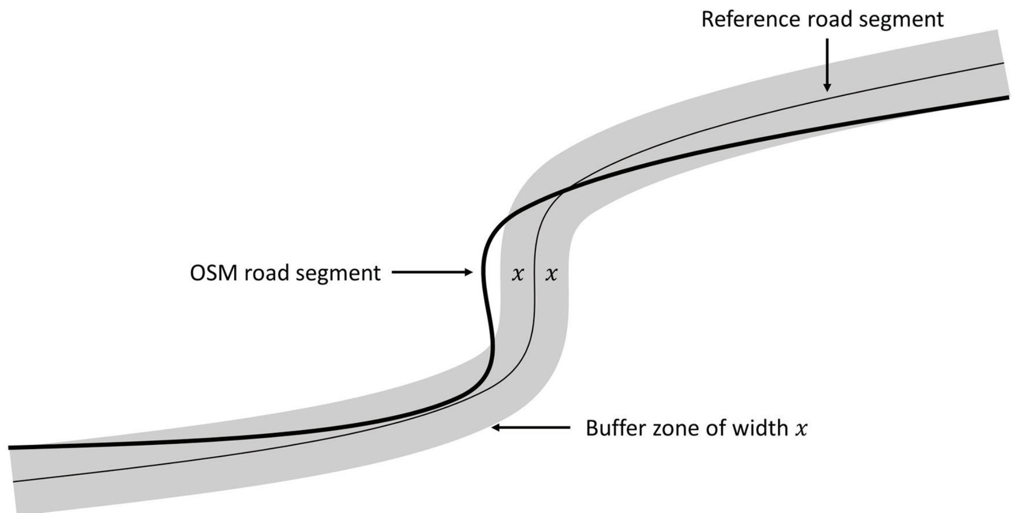


Table 5. Matches of attributes

Field Name	Field Type	Field Description
name	Text	Full street name
length	Number	Length of the road segment
rank	Number	New road classifications
UID	Number	Unique ID
preDir	Text	Prefix direction
preType	Text	Prefix street type
stName	Text	Street name component
sufType	Text	Suffix street type
sufDir	Text	Suffix direction
tunnel	Number	1 = tunnel; 0 = not tunnel
bridge	Number	1 = bridge; 0 = not bridge
oneway	Number	1 = oneway; 0 = two ways; -1 = incorrect input

Table 6. Matches of road classifications

New Rank	DMTI Road Types	OSM Road Types	
0	N.A.	Unclassified	
1	Expressways	Motorway	Motorway_ Link
2	Primary Highways	Trunk	Trunk_Link
3	Secondary Highways	Primary	Primary_Link
4	Major Roads	Secondary	Secondary_ Link
5	Local Roads	Tertiary	Tertiary_Link
		Residential	Service
6	Trails	Footway	Steps
	Proposed Roads	Path	Track
		Raceway	Cycleway

Positional Accuracy

To improve the results of the geometric feature matching, rank 6 is excluded from the following analysis. Figure 8 shows the proportions of OSM road segments that fall within the buffers of DMTI road segments with a range from 1 to 10 m. Approximately all ranks of roads have a logarithmic increase of their positional accuracy. The average positional offset is 2.3 m, which is significantly better than the results in London, UK and England in 2007 (5.8 m) (Haklay, 2010) and 2009 (7.9 m) (Antoniou, 2011).

Figure 6. Classified road lengths in London, Canada

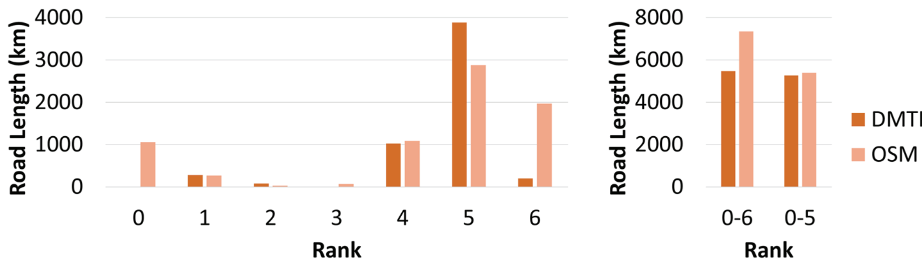
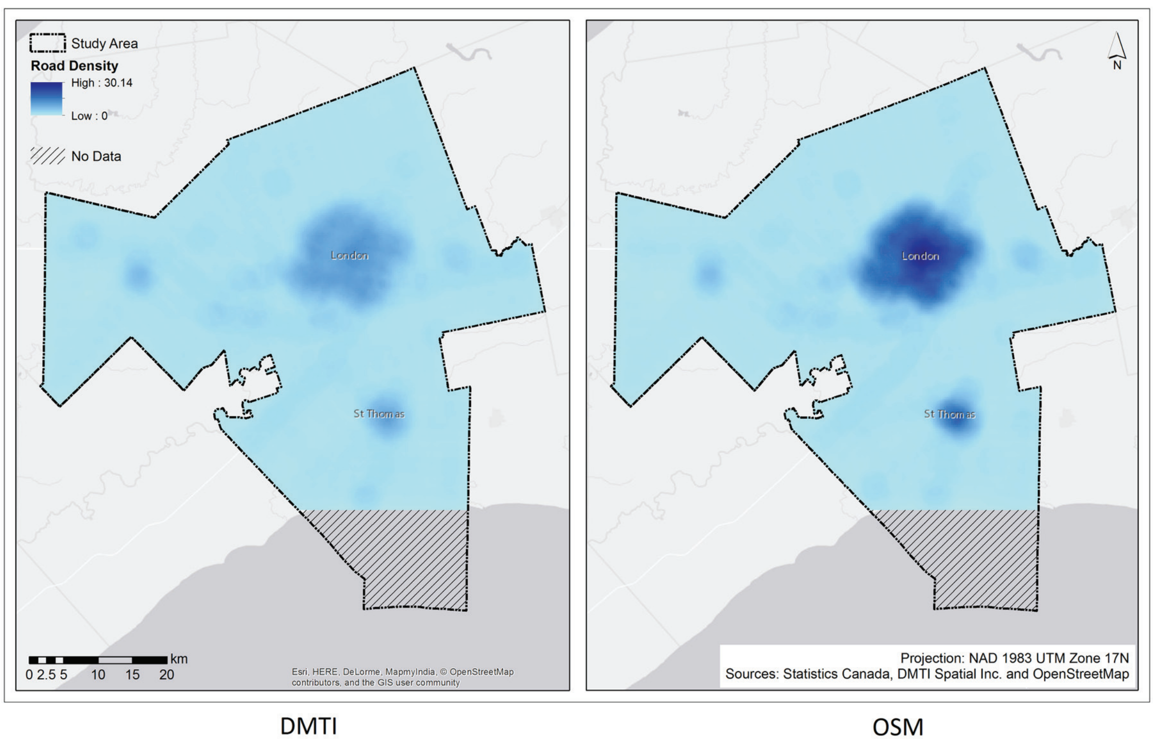
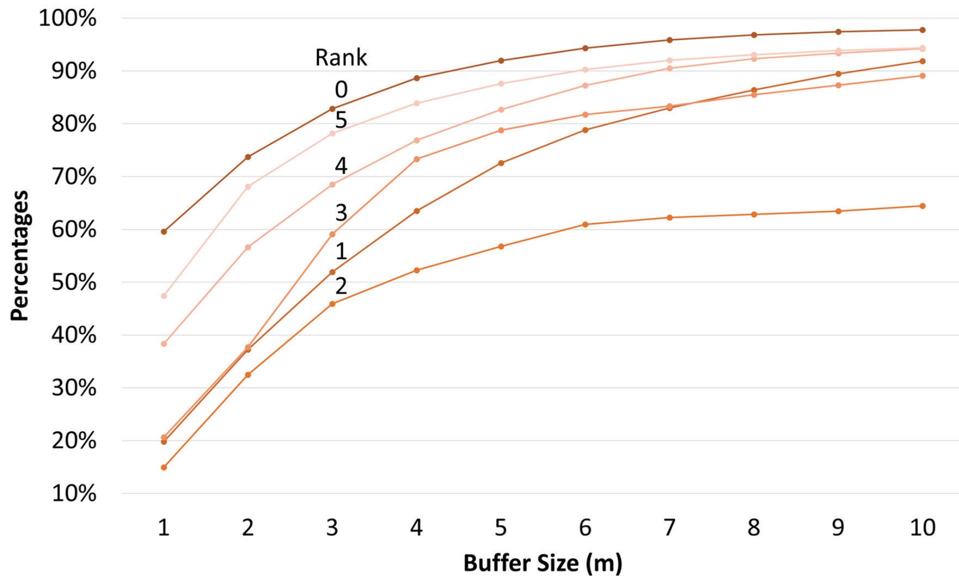


Figure 7. Road density (m/km²) in London, Canada



At buffer size of 1 m, the positional accuracy ranges from 14.9 to 59.6%. The accuracy increases at a relatively fast rate until 6 m. After that, the accuracy starts to only increase gently. Over 86% of road segments have positional errors within 5 m, which is also better than 73% of road segments in Germany in 2009(Ludwig et al., 2011). At buffer size of 10 m, most ranks have over 91% of positional accuracy except rank 2 and 3. However, the lengths of roads in these two ranks of exception are relatively short (Table 4), which means their results may not be representative. The most accurate rank at the 10-m buffer is rank 0 (local roads in suburban areas).

Figure 8. Trends of the OSM positional accuracy in London, Canada



Attribute Accuracy

The percentages of attribute accuracy are calculated by road lengths as well. Table 7 lists the proportions of presented OSM tags against the available DMTI attributes. The numeric fields are not included since all OSM road segments have a rank (rank 0 = unclassified) and the remainders have limited number of entries. The presence rates are mostly very high except for sufDir (e.g., N, S, W, E), which probably indicates that the suffix directions are not the primary concerns to the OSM users or not well-known to the OSM contributors. The presence rate of rank 1 under sufType is extremely low as well, which is because a large number of highway segments miss the suffix type “RAMP”. The overall rate of sufType is not affected because of the relatively short length of highway. The attribute completeness of London, Canada are actually superior comparing to French streets (85% for types and 43% for names) (Girres & Touya, 2010) and German streets (82.5% to 94.4% for names) (Ludwig et al., 2011) in 2009.

Table 8 presents the absolute difference of the numeric attributes between the OSM and DMTI data. Only 70.6% of the OSM road segments have matched road classifications, which is largely due to the unclassified local roads in suburban area (the 21.1% that have a difference of 5). The rest of the fields have almost perfect accuracy; however, the results need to be interpreted with caution because of the short total length of tunnels, bridges and oneway roads. Still, the nearly 98% of oneway flag accuracy in London, Canada is better than the 16% completeness in France in 2009 (Girres & Touya, 2010).

Table 9 lists the Levenshtein distance of the text fields. Overall, the longer the field content, the larger the Levenshtein distance. Therefore, preDir and sufDir have excellent accuracy since the length of these fields is one letter. Another reason of the nearly perfect accuracy of preDir is due to its small number of entries, and so does preType. The accuracy results of stName and sufType are lower than the others, but

Table 7. Tag presence of the text fields

OSM Rank	Percent	OSM Rank	Percent
preDir		sufType	
4	100.0%	0	99.5%
5	91.6%	1	27.9%
Overall	93.2%	2	92.8%
preType		3	79.0%
1	99.7%	4	98.5%
3	94.1%	5	97.2%
Overall	99.5%	Overall	96.8%
stName		sufDir	
0	99.6%	0	0.0%
1	97.7%	1	42.0%
2	100.0%	4	69.3%
3	100.0%	5	46.8%
4	99.5%	-	-
5	97.4%	-	-
Overall	98.4%	Overall	62.1%

Table 8. Number difference of the numeric fields

Difference	Percent	Difference	Percent
rank		bridge	
0	70.6%	0	99.5%
1	7.1%	1	0.5%
2	0.9%	oneway	
3	0.0%	0	97.9%
4	0.2%	1	2.1%
5	21.1%	2	0.0%
tunnel		-	-
0	100.0%	-	-
1	0.0%	-	-

Table 9. Levenshtein distance (LD) of the text fields

LD	Percent	LD	Percent
preDir		sufDir	
0	99.7%	0	97.3%
1	0.3%	1	2.7%
preType		sufType	
0	99.4%	0	89.2%
3	0.6%	1	0.0%
stName		2	4.6%
0	86.1%	3	2.2%
1 to 3	3.1%	4	3.8%
> 3	10.9%	5	0.2%

still above 85%. A Levenshtein distance of 1 to 3 usually represents spelling mistakes (Girres & Touya, 2010). However, a small portion of stName and sufType have large Levenshtein distance that is greater than 3. The large Levenshtein distances do not affect the overall accuracy as the average Levenshtein distance of stName is only 0.8, which is significantly smaller than the same variable (4.96) of lake names in France in 2009 (Girres & Touya, 2010).

CONCLUSION

Although OSM has better data completeness and overall good positional and attribute accuracy comparing to DMTI, it still has some quality issues. For example, the majority of local roads in rural areas remain unclassified. Misspelling of street names and suffix types still exists, and a large number of suffix directions are missing as well. Still, the general OSM quality of London, Canada in 2016 has greatly improved comparing to previous studies of US and European regions. An interesting finding is that the local roads in suburban areas (rank 0) actually have the highest level of positional accuracy, which violates the assumption brought up at the beginning of the case study section. This high accuracy of local roads in rural areas is perhaps due to the data import from an old version of NRN starting in 2008 ("Canada Import Status - OpenStreetMap Wiki," 2015) and the limited user-editing afterwards. Hence, it is worth to explore the OSM quality at a larger scale. For instance, there are no reference roads classified as secondary highways (rank 3) in the London CMA, which will not be a problem once the study area is expanded to the national level. In addition, an exploration is still needed for evaluating the trail data (rank 6) if a reference dataset is available. Other future research questions pertaining to OSM and VGI are as follows:

- Which data source, the commercial organization, the governmental data bureau or VGI, should be used under which circumstances?
- Are there better and more efficient methods to evaluate the extrinsic (when a reference dataset is available) and intrinsic (e.g., data history analysis) OSM quality?
- How can one accurately automate the quality assessment process?
- How can one improve OSM quality in general?

REFERENCES

Advertise with us - Gasbuddy Gas Prices. (n.d.). Retrieved July 3, 2016, from https://blog.gasbuddy.com/advertising.aspx

Amelunxen, C. (2010). On the suitability of Volunteered Geographic Information for the purpose of geocoding. *GI_Forum*, 1–8.

Antoniou, V. (2011). *User Generated Spatial Content: An Analysis of the Phenomenon and its Challenges for Mapping Agencies*. Retrieved from http://discovery.ucl.ac.uk/1318053/

Ballatore, A., & Zipf, A. (2015). A conceptual quality framework for volunteered geographic information. *12th International Conference on Spatial Information Theory, COSIT 2015*. Springer Verlag. doi:10.1007/978-3-319-23374-1_5

Barron, C., Neis, P., & Zipf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, *18*(6), 877–895. doi:10.1111/tgis.12073

Basiri, A., Jackson, M., Amirian, P., Pourabdollah, A., Sester, M., Winstanley, A., … Zhang, L. (2016). Quality assessment of OpenStreetMap data using trajectory mining. *Geo-Spatial Information Science, 19*(1), 56–68. 10.1080/10095020.2016.1151213

Brovelli, M. A., Minghini, M., Molinari, M., & Mooney, P. (2016). *Towards an Automated Comparison of OpenStreetMap with Authoritative Road Datasets. Transactions in GIS*. doi:10.1111/tgis.12182

Buehler, R., & Pucher, J. (2012). Cycling to work in 90 large American cities: New evidence on the role of bike paths and lanes. *Transportation*, *39*(2), 409–432. doi:10.1007/s11116-011-9355-8

Camboim, S. P., Meza Bravo, J. V, & Sluter, C. R. (2015). An investigation into the completeness of, and the updates to, OpenStreetMap data in a heterogeneous area in Brazil. *ISPRS International Journal of Geo-Information, 4*(3), 1366–1388. 10.3390/ijgi4031366

Canada Import Status - OpenStreetMap Wiki. (2015). Retrieved July 3, 2016, from http://wiki.openstreetmap.org/wiki/Canada_Import_Status

Canavosio-Zuzelski, R., Agouris, P., & Doucette, P. (2013). *A Photogrammetric Approach for Assessing Positional Accuracy of OpenStreetMap© Roads. International Journal of Geo-Information, 2(2)*.

CBC News. (2013). *CBC News - Google buys Waze mapping app*. Retrieved July 3, 2016, from http://www.cbc.ca/m/touch/business/story/1.1306058

Cipełuch, B., Jacob, R., Mooney, P., & Winstanley, A. C. (2010). Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. In *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences* (p. 337). CONF, University of Leicester.

Coleman, D. J. (2013). Potential Contributions and Challenges of VGI for Conventional Topographic Base-Mapping Programs. In D. Sui, S. Elwood, & M. Goodchild (Eds.), Crowdsourcing Geographic Knowledge SE - 14 (pp. 245–263). CHAP, Springer Netherlands. doi:10.1007/978-94-007-4587-2_14

Coleman, D. J., Georgiadou, Y., & Labonte, J. (2009). Volunteered geographic information: The nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research*, *4*(1), 332–358.

Corcoran, P., Mooney, P., & Bertolotto, M. (2013). Analysing the growth of OpenStreetMap networks. *Spatial Statistics*, *3*, 21–32. doi:10.1016/j.spasta.2013.01.002

DMTI Spatial Inc. (2015). *CanMap® Content Suite Support Manual*. Markham.

Dobson, M. (2013). VGI as a Compilation Tool for Navigation Map Databases. In D. Sui, S. Elwood, & M. Goodchild (Eds.), Crowdsourcing Geographic Knowledge SE - 17 (pp. 307–327). CHAP, Springer Netherlands. doi:10.1007/978-94-007-4587-2_17

Dorn, H., Törnros, T., & Zipf, A. (2015). Quality evaluation of VGI using authoritative data-a comparison with land use data in southern Germany. *ISPRS International Journal of Geo-Information, 4*(3), 1657–167110.3390/ijgi4031657

Eckle, M., & De Albuquerque, J. P. (2015). Quality assessment of remote mapping in OpenStreetMap for disaster management purposes. In L. A. Palen, T. Comes, M. Buscher, A. L. Hughes, & L. A. Palen (Eds.), *12th International Conference on Information Systems for Crisis Response and Management, ISCRAM 2015*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84947764602&partnerID=40&md5=772a654ed328051eca2e2591b6430aa9

El-Ashmawy, K. L. A. (2016). Testing the positional accuracy of OpenStreetMap data for mapping applications. *Geodesy and Cartography, 42*(1), 25–30. 10.3846/20296991.2016.1160493

Elwood, S. (2006). Critical issues in participatory GIS: Deconstructions, reconstructions, and new research directions. *Transactions in GIS*, *10*(5), 693–708. doi:10.1111/j.1467-9671.2006.01023.x

Elwood, S., Goodchild, M., & Sui, D. (2013). Prospects for VGI Research and the Emerging Fourth Paradigm. In D. Sui, S. Elwood, & M. Goodchild (Eds.), Crowdsourcing Geographic Knowledge SE - 20 (pp. 361–375). CHAP, Springer Netherlands. doi:10.1007/978-94-007-4587-2_20

Estima, J., & Painho, M. (2013). Exploratory analysis of OpenStreetMap for land use classification. *Geocrowd*, *13*, 39–46. doi:10.1145/2534732.2534734

Estima, J., & Painho, M. (2015). Investigating the Potential of OpenStreetMap for Land Use/Land Cover Production: A Case Study for Continental Portugal. In J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience SE - 14 (pp. 273–293). CHAP, Springer International Publishing. doi:10.1007/978-3-319-14280-7_14

Fan, H., Yang, B., Zipf, A., & Rousell, A. (2015). A polygon-based approach for matching OpenStreetMap road networks with regional transit authority data. *International Journal of Geographical Information Science, 30*, 1–17. 10.1080/13658816.2015.1100732

Fan, H., Zipf, A., Fu, Q., & Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, *0*(0), 1–20. doi:10.1080/13658816.2013.867495

Feick, R., & Roche, S. (2013). Understanding the Value of VGI. In D. Sui, S. Elwood, & M. Goodchild (Eds.), Crowdsourcing Geographic Knowledge SE - 2 (pp. 15–29). CHAP, Springer Netherlands doi:10.1007/978-94-007-4587-2_2

Forghani, M., & Delavar, M. (2014). A Quality Study of the OpenStreetMap Dataset for Tehran. *ISPRS International Journal of Geo-Information*, *3*(2), 750–763. doi:10.3390/ijgi3020750

Friedman, A., & Kohler, B. (2003). Bidimensional regression: Assessing the configural similarity and accuracy of cognitive maps and other two-dimensional data sets. *Psychological Methods*, *8*(4), 468–491. doi:10.1037/1082-989X.8.4.468 PMID:14664683

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, *24*(3), 189–206. doi:10.1111/j.1538-4632.1992.tb00261.x

Girres, J. F., & Touya, G. (2010). Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, *14*(4), 435–459. doi:10.1111/j.1467-9671.2010.01203.x

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, *69*(4), 211–221. doi:10.1007/s10708-007-9111-y

Goodchild, M. F. (2009). What problem? Spatial autocorrelation and geographic information science. *Geographical Analysis*, *41*(4), 411–417. doi:10.1111/j.1538-4632.2009.00769.x

Goodchild, M. F. (2013). The quality of big (geo)data. *Dialogues in Human Geography*, *3*(3), 280–284. doi:10.1177/2043820613513392

Goodchild, M. F., & Hunter, G. J. (1997). A simple positional accuracy measure for linear features. *International Journal of Geographical Information Science, 11*(3), 299–306. doi:10.1080/136588197242419

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, *1*, 110–120. doi:10.1016/j.spasta.2012.03.002

Graser, A., Straub, M., & Dragaschnig, M. (2014). Towards an open source analysis toolbox for street network comparison: Indicators, tools and results of a comparison of osm and the official Austrian reference graph. *Transactions in GIS*, *18*(4), 510–526. doi:10.1111/tgis.12061

Gröchenig, S., Brunauer, R., & Rehrl, K. (2014). Estimating completeness of VGI datasets by analyzing community activity over time periods. In C. Granell, S. Schade, & J. Huerta (Eds.), *17th AGILE Conference on Geographic Information Science, AGILE 2014* (pp. 3–18). Kluwer Academic Publishers. doi:10.1007/978-3-319-03611-3_1

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. *Environment and Planning. B, Planning & Design*, *37*(4), 682–703. doi:10.1068/b35097

Haklay, M. (2013). Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. In D. Sui, S. Elwood, & M. Goodchild (Eds.), Crowdsourcing Geographic Knowledge SE - 7 (pp. 105–122). CHAP, Springer Netherlands. doi:10.1007/978-94-007-4587-2_7

Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus Law to Volunteered Geographic Information. *The Cartographic Journal*, *47*(4), 315–322. doi:10.1179/000870410X12911304958827

Haklay, M., & Weber, P. (2008). OpenStreet map: User-generated street maps. *IEEE Pervasive Computing / IEEE Computer Society [and] IEEE Communications Society*, *7*(4), 12–18. doi:10.1109/MPRV.2008.80

Hashemi, P., & Ali Abbaspour, R. (2015). Assessment of Logical Consistency in OpenStreetMap Based on the Spatial Similarity Concept. In J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience SE - 2 (pp. 19–36). CHAP, Springer International Publishing. doi:10.1007/978-3-319-14280-7_2

Hayakawa, T., Imi, Y., & Ito, T. (2012). Analysis of quality of data in OpenStreetMap. In *2012 IEEE 14th International Conference on Commerce and Enterprise Computing, CEC 2012* (pp. 131–134). doi:10.1109/CEC.2012.29

Hecht, R., Kunze, C., & Hahmann, S. (2013). Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. *International Journal of Geo-Information, 2*, 1066–1091. 10.3390/ijgi2041066

Helbich, M., Amelunxen, C., & Neis, P. (2012). Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata.*Proceedings of GI_Forum 2012: Geovisualization, Society and Learning*, 24–33. doi:10.1080/13658816

Hochmair, H. H., & Zielstra, D. (2013). Development and Completeness of Points Of Interest in Free and Proprietary Data Sets: A Florida Case Study. *Gi_Forum 2013: Creating the Gisociety*, 39–48. 10.1553/giscience2013s39

Hochmair, H. H., Zielstra, D., & Neis, P. (2015). Assessing the Completeness of Bicycle Trail and Designated Lane Features in OpenStreetMap for the United States and Europe. *Transactions in GIS*, *19*(1), 63–81. doi:10.1111/tgis.12081

Hunter, G. J. (1999). New tools for handling spatial data quality: Moving from academic concepts to practical reality. *URISA Journal*, *11*(2), 25–34.

Index of /planet/full-history. (2016). Retrieved July 4, 2016, from http://planet.openstreetmap.org/planet/full-history/

IRC - OpenStreetMap Wiki. (2015). Retrieved July 4, 2016, from http://wiki.openstreetmap.org/wiki/IRC

ISO, T. C. 211. (2002). *ISO 19113, Geographic Information–Quality Principles.* The International Organization for Standardisation.

Jackson, S., Mullen, W., Agouris, P., Crooks, A., Croitoru, A., & Stefanidis, A. (2013). Assessing Completeness and Spatial Error of Features in Volunteered Geographic Information. *ISPRS International Journal of Geo-Information*, *2*(2), 507–530. doi:10.3390/ijgi2020507

Jiang, B. (2013). Volunteered Geographic Information and Computational Geography: New Perspectives. In D. Sui, S. Elwood, & M. Goodchild (Eds.), Crowdsourcing Geographic Knowledge SE - 8 (pp. 125–138). CHAP, Springer Netherlands. doi:10.1007/978-94-007-4587-2_8

Jilani, M., Corcoran, P., & Bertolotto, M. (2014). Automated highway tag assessment of openstreetmap road networks. In M. Schneider, M. Gertz, Y. Huang, J. Sankaranarayanan, & J. Krumm (Eds.), *22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM SIGSPATIAL GIS 2014* (*Vol. 04–07*–Nove, pp. 449–452). Association for Computing Machinery. doi:10.1145/2666310.2666476

Jokar Arsanjani, J., & Bakillah, M. (2014). Understanding the potential relationship between the socio-economic variables and contributions to OpenStreetMap. *International Journal of Digital Earth*, *8*(10), 1–16. doi:10.1080/17538947.2014.951081

Jokar Arsanjani, J., Barron, C., Bakillah, M., & Helbich, M. (2013). Assessing the Quality of OpenStreetMap Contributors together with their Contributions.*16th AGILE Conference*.

Jokar Arsanjani, J., Helbich, M., Bakillah, M., Hagenauer, J., & Zipf, A. (2013). Toward mapping land-use patterns from volunteered geographic information. *International Journal of Geographical Information Science*, *27*(12), 2264–2278. doi:10.1080/13658816.2013.800871

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

Jokar Arsanjani, J., Mooney, P., Helbich, M., & Zipf, A. (2015). An Exploration of Future Patterns of the Contributions to OpenStreetMap and Development of a Contribution Index. *Transactions in GIS, 19*(6), 896–914.10.1111/tgis.12139

Jokar Arsanjani, J., Mooney, P., Zipf, A., & Schauss, A. (2015). Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets. In J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience SE - 3 (pp. 37–58). CHAP, Springer International Publishing.doi:10.1007/978-3-319-14280-7_3

Jokar Arsanjani, J., & Vaz, E. (2015). An assessment of a collaborative mapping approach for exploring land use patterns for several European metropolises. *International Journal of Applied Earth Observation and Geoinformation*, *35*, 329–337. doi:10.1016/j.jag.2014.09.009

Jokar Arsanjani, J., Zipf, A., Mooney, P., & Helbich, M. (2015). An Introduction to OpenStreetMap in Geographic Information Science: Experiences, Research, and Applications. In J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience SE - 1 (pp. 1–15). CHAP, Springer International Publishing.doi:10.1007/978-3-319-14280-7_1

Kalantari, M., & La, V. (2015). Assessing OpenStreetMap as an Open Property Map. In J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience SE - 13 (pp. 255–272). CHAP, Springer International Publishing.doi:10.1007/978-3-319-14280-7_13

Keßler, C., Theodore, R., & De Groot, A. (2013). Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of OpenStreetMap. In *Geographic Information Science at the Heart of Europe* (pp. 21–37). Springer; doi:10.1007/978-3-319-00615-4_2

Koukoletsos, T., Haklay, M., & Ellul, C. (2012). Assessing Data Completeness of VGI through an Automated Matching Procedure for Linear Data. *Transactions in GIS*, *16*(4), 477–498. doi:10.1111/j.1467-9671.2012.01304.x

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics, Doklady*, *10*, 707.

Lin, W. (2013). When Web 2.0 Meets Public Participation GIS (PPGIS): VGI and Spaces of Participatory Mapping in China. In D. Sui, S. Elwood, & M. Goodchild (Eds.), Crowdsourcing Geographic Knowledge SE - 6 (pp. 83–103). CHAP, Springer Netherlands.doi:10.1007/978-94-007-4587-2_6

Ludwig, I., Voss, A., & Krause-traudes, M. (2011). A Comparison of the Street Networks of Navteq and OSM in Germany. In *Advancing Geoinformation Science for a Changing World* (Vol. 1, pp. 65–84). Berlin: Springer-Verlag; doi:10.1007/978-3-642-19789-5_4

Mailing lists - OpenStreetMap Wiki. (2016). Retrieved July 4, 2016, from http://wiki.openstreetmap.org/wiki/Mailing_lists

Mashhadi, A., Quattrone, G., & Capra, L. (2015). The Impact of Society on Volunteered Geographic Information: The Case of OpenStreetMap. In J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience SE - 7 (pp. 125–141). CHAP, Springer International Publishing. doi:10.1007/978-3-319-14280-7_7

McMaster, R. B. (1986). A statistical analysis of mathematical measures for linear simplification. *The American Cartographr*, *13*(2), 103–116. doi:10.1559/152304086783900059

Meier, J. C. H. (2015). *An Analysis of Quality for Volunteered Geographic Information*. JOUR.

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal, 80*(4), 449–461.10.1007/s10708-014-9602-6

Mohammadi, N., & Malek, M. (2015). Artificial intelligence-based solution to estimate the spatial accuracy of volunteered geographic data. *Journal of Spatial Science, 60*(1), 119–135.10.1080/14498596.2014.927337

Mondzech, J., & Sester, M. (2011). Quality analysis of OpenStreetMap data based on application needs. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *46*(2), 115–125. doi:10.3138/carto.46.2.115

Mooney, P., & Corcoran, P. (2012a). Characteristics of Heavily Edited Objects in OpenStreetMap. Future Internet, 4(4), 285–305.doi:10.3390/fi4010285

Mooney, P., & Corcoran, P. (2012b). The Annotation Process in OpenStreetMap. *Transactions in GIS, 16*(4), 561–579. Journal Article.10.1111/j.1467-9671.2012.01306.x

Mooney, P., & Corcoran, P. (2014). Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. *Transactions in GIS*, *18*(5), 633–659. doi:10.1111/tgis.12051

Mooney, P., Corcoran, P., & Winstanley, A. C. (2010). Towards quality metrics for OpenStreetMap. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems GIS 10*, 514–517.doi:10.1145/1869790.1869875

Moovit Company Overview. (2014). Retrieved July 3, 2016, from http://moovitapp.com/wp-content/uploads/2014/07/Moovit-At-A-Glance-USA-4-14.pdf

Mullen, W. F., Jackson, S. P., Croitoru, A., Crooks, A., Stefanidis, A., & Agouris, P. (2015). Assessing the impact of demographic characteristics on spatial error in volunteered geographic information features. *GeoJournal, 80*(4), 587–605.10.1007/s10708-014-9564-8

Natural Resources Canada. (2015). *National Road Network (NRN) - Full metadata*. Retrieved April 23, 2016, from http://geogratis.gc.ca/api/en/nrcan-rncan/ess-sst/c0d1f299-179c-47b2-bcd8-da1ba68a8032.html#full-screen-meta

Neis, P., & Zielstra, D. (2014). Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet*, *6*(1), 76–106. doi:10.3390/fi6010076

Neis, P., Zielstra, D., & Zipf, A. (2011). The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. *Future Internet*, *4*(4), 1–21. doi:10.3390/fi4010001

Neis, P., & Zipf, A. (2012). Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *1*(3), 146–165. doi:10.3390/ijgi1020146

OSMstats - Statistics of the free wiki world map. (2016). Retrieved July 3, 2016, from http://osmstats.neis-one.org/

Parr, D. A. (2015). *The production of volunteered geographic information: A study of OpenStreetMap in the United States*. Academic Press.

Pourabdollah, A., Morley, J., Feldman, S., & Jackson, M. (2013). Towards an Authoritative OpenStreetMap: Conflating OSM and OS OpenData National Maps Road Network. *ISPRS International Journal of Geo-Information*, *2*(3), 704–728. doi:10.3390/ijgi2030704

Quality assurance - OpenStreetMap Wiki. (2016). Retrieved July 3, 2016, from http://wiki.openstreetmap.org/wiki/Quality_assurance

Quattrone, G., Mashhadi, A., Quercia, D., Smith-Clarke, C., & Capra, L. (2014). Modelling growth of urban crowd-sourced information. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining - WSDM '14*, 563–572.doi:10.1145/2556195.2556244

Raymond, E. S. (2001). *The Cathedral & the Bazaar: Musings on linux and open source by an accidental revolutionary*. O'Reilly Media, Inc.

Sehra, S. S., Singh, J., & Rai, H. S. (2015). Quality assessment of crowdsourced data against custom recorded map data. *Indian Journal of Science and Technology, 8*(33).10.17485/ijst/2015/v8i33/79884

Siebritz, L., Sithole, G., & Zlatanova, S. (2012). Assessment of the homogeneity of volunteered geographic information in South Africa. In M. Shortis & M. Madden (Eds.), *22nd Congress of the International Society for Photogrammetry and Remote Sensing, ISPRS 2012* (Vol. 39, pp. 553–558). International Society for Photogrammetry and Remote Sensing. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84901710768&partnerID=40&md5=e65dd4e6feb4732bcd9d94f1c39a4e19

Statistics Canada. (2012). Focus on Geography Series, 2011 Census. Statistics Canada Catalogue no. 98-310-XWE2011004. Ottawa: Author.

Stein, K., Kremer, D., & Schlieder, C. (2015). Spatial Collaboration Networks of OpenStreetMap. In J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience SE - 9 (pp. 167–186). CHAP, Springer International Publishing.doi:10.1007/978-3-319-14280-7_9

Tenney, M. (2014). *Quality Evaluations on Canadian OpenStreetMap Data*. Retrieved from http://rose.geog.mcgill.ca/ski/system/files/fm/2014/tenney.pdf

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234–240. doi:10.2307/143141

Tobler, W. R. (1994). Bidimensional regression. *Geographical Analysis*, *26*(3), 187–212. doi:10.1111/j.1538-4632.1994.tb00320.x

Touya, G., & Brando-Escobar, C. (2013). Detecting Level-of-Detail Inconsistencies in Volunteered Geographic Information Data Sets. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *48*(2), 134–143. doi:10.3138/carto.48.2.1836

Touya, G., & Reimer, A. (2015). Inferring the Scale of OpenStreetMap Features. In J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience SE - 5 (pp. 81–99). CHAP, Springer International Publishing.doi:10.1007/978-3-319-14280-7_5

Van Exel, M., Dias, E., & Fruijtier, S. (2010). The impact of crowdsourcing on spatial data quality indicators. *Proceedings of GiScience*, *2011*, 1–4.

Van Oort, P. (2006). *Spatial data quality: from description to application. Publications on Geodesy 60*. DISS, WUR Wageningen UR. Retrieved from http://library.wur.nl/WebQuery/wdab/1788022

Vandecasteele, A., & Devillers, R. (2015). Improving Volunteered Geographic Information Quality Using a Tag Recommender System: The Case of OpenStreetMap. In J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience SE - 4 (pp. 59–80). CHAP, Springer International Publishing.doi:10.1007/978-3-319-14280-7_4

Vauglin, F. (1997). *Modèles statistiques des imprécisions géométriques des objets géographiques linéaires.* DISS.

Vaz, E., & Jokar Arsanjani, J. (2015). Crowdsourced mapping of land use in urban dense environments: An assessment of Toronto. *Canadian Geographer, 59*(2), 246–255.10.1111/cag.12170

Wang, M., Li, Q., Hu, Q., & Zhou, M. (2013). Quality analysis of open street map data. In E. Guilbert, B. Wu, & J. Shi (Eds.), *8th International Symposium on Spatial Data Quality* (2W1, Vol. 40, pp. 155–158). International Society for Photogrammetry and Remote Sensing. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84929074534&partnerID=40&md5=cc6cb4934bc6f9cdeea8a05303363311

Yang, A., Fan, H., & Jing, N. (2016). *Amateur or professional: Assessing the expertise of major contributors in openstreetmap based on contributing behaviors. ISPRS International Journal of Geo-Information, 5(2)*. doi:10.3390/ijgi5020021

Zhang, X., & Ai, T. (2015). How to model roads in openstreetmap? A method for evaluating the fitness-for-use of the network for navigation. In Y. Leung & F. Harvey (Eds.), *16th International Symposium on Spatial Data Handling, SDH 2014* (*Vol. 19*, pp. 143–162). Springer Berlin.doi:10.1007/978-3-319-19950-4_9

Zhao, Y. J., Zhou, X. G., Li, G. Q., & Xing, H. F. (2016). *A Spatio-Temporal VGI Model Considering Trust-Related Information. ISPRS International Journal of Geo-Information, 5(2), 19*. doi:10.3390/ijgi5020010

Zhou, P., Huang, W., & Jang, J. (2014). Validation analysis of OpenStreetMap data in some areas of China. In J. Jiang & H. Zhang (Eds.), *ISPRS Technical Commission IV Symposium 2014* (4th ed.; *Vol. 40*, pp. 383–391). International Society for Photogrammetry and Remote Sensing.doi:10.5194/isprsarchives-XL-4-383-2014

Zielstra, D., Hochmair, H. H., & Neis, P. (2013). Assessing the effect of data imports on the completeness of openstreetmap - A United States case study. *Transactions in GIS*, *17*(3), 315–334. doi:10.1111/tgis.12037

Zielstra, D., & Zipf, A. (2010). A comparative study of proprietary geodata and volunteered geographic information for Germany.*13th AGILE International Conference on Geographic Information Science 2010 Guimarães, Portugal, 1*, 1–15.

KEY TERMS AND DEFINITIONS

Attribute Accuracy: A measure of the correctness of attributes associated with geometrical shapes. It is also known as thematic accuracy.

Authoritative Dataset: A validated dataset published by governmental agencies or commercial organizations.

Buffer Analysis: A method of measuring positional accuracy. Buffers are created around objects from an authoritative dataset, and the percentages of objects from a test dataset that fall within the buffers are calculated.

Completeness: A measure of the comprehensiveness of a dataset. Data completeness not only reports how much data is missing, but also the amount of data that should be excluded.

Levenshtein Distance: A numerical value that reflects the differences of two strings, which is also known as edit distance.

Linus' Law: Errors can be easily identified with enough number of users in a software development environment.

OpenStreetMap: A crowd-sourced online platform that allows citizen scientists to edit maps of their neighborhood and the world using satellite images.

Positional Accuracy: A measure of the relative and absolute accuracy of coordinate values.

ENDNOTES

1 See https://mapzen.com/data/metro-extracts/

2 See https://mapzen.com/documentation/metro-extracts/overview/#choose-a-file-format

3 See http://geo2.scholarsportal.info/

APPENDIX

Table 10. Summary of recent literatures on quality analysis of OSM

Studies	Time	Study Areas	Reference Data Sources	Data Types			Quality Criteria or Methodology						
				POI	Line	Poly	C	PA	AA	LC	SA	TQ	L
Amelunxen (2010)	N/A	North Rhine-Westfalia, Germany	Geocoding service by Google	x				x					
Cipełuch, Jacob, Mooney, & Winstanley (2010)	2010	Ireland	Google Maps and Bing Maps		x		x		x	x			
Girres & Touya (2010)	2009	France	BD TOPO	x	x	x	x	x	x	x	x	x	x
Haklay (2010)	2007	England, UK	OS Meridian 2		x		x	x					
Haklay, Basiouka, Antoniou, & Ather (2010)	2007	London and England, UK	OS Meridian 2		x		Relationship between average positional error and number of contributors						
Mooney, Corcoran, & Winstanley (2010)	2010	European regions	N/A		x	x			x	x			
Zielstra & Zipf (2010)	2009	Germany^	Tele Atlas		x		x					x	
Antoniou (2011)	2009	England, UK	OS Meridian 2		x			x		x			
Ludwig, Voss, & Krause-traudes (2011)	2009	Germany^	Navteq		x		x	x	x				
Mondzech & Sester (2011)	N/A	Germany	ATKIS		x		Accessibility and length of simulated routes						
Neis, Zielstra, & Zipf (2011)	2007 to 2011	Germany^	TomTom		x		x			x		x	
Hayakawa, Imi, & Ito (2012)	2012	Japan and other regions	N/A	x	x	x	x						
Helbich, Amelunxen, & Neis (2012)	N/A	Germany	Tele Atlas		x			x					
Koukoletsos, Haklay, & Ellul (2012)	N/A	London and Newcastle, UK	OS ITN layer of MasterMap		x		x						
Mooney & Corcoran (2012a)	2011	UK and Ireland	N/A (User behavior)		x		Correlation between numbers of contributors and numbers of tags						
Mooney & Corcoran (2012b)	2011	UK, Ireland, Germany and Austria	N/A		x	x			x				
Siebritz et al. (2012)	2006 to 2011	South Africa	NMA	x	x							x	
Canavosio-Zuzelski, Agouris, & Doucette (2013)	2011	Purdue University, US	USGS National Map and TIGER/Line		x			x					
Corcoran, Mooney, & Bertolotto (2013)	2007 to 2011	Ireland	N/A		x							x	
Estima & Painho (2013, 2015)	2013	Portugal^	CLC	x					x				
Hecht, Kunze, & Hahmann (2013)	2011, 2012	Germany	Official building polygon dataset and ATKIS			x	x					x	
Hochmair & Zielstra (2013)	2012	Florida, US	TomTom, NAVTEQ, ESRI and TIGER/Line	x			x					x	
Jackson et al. (2013)	2011	Denver, US	ORNL	x			x	x					
Jokar Arsanjani, Barron, Bakillah, & Helbich (2013)	2012	Heidelberg, Germany	BKG		x		x	x			x		

continued on following page

Table 10. Continued

Studies	Time	Study Areas	Reference Data Sources	Data Types			Quality Criteria or Methodology						
				POI	Line	Poly	C	PA	AA	LC	SA	TQ	L
Jokar Arsanjani, Helbich, Bakillah, Hagenauer, & Zipf (2013)	2012	Vienna, Austria	GMESUA			x			x				
Keßler, Theodore, & Groot (2013)	2011	Münster, Germany	N/A (Data trust and vandalism)	x	x	x	Trustworthiness (e.g., versions, users, confirmations and tag corrections)						
Pourabdollah, Morley, Feldman, & Jackson (2013)	N/A	UK^	OS VMD		x				x				
Touya & Brando-Escobar (2013)	N/A	France	N/A	x	x	x	Level of Details						
Wang, Li, Hu, & Zhou (2013)	N/A	Wuhan, China	NavInfo		x		x	x	x				
Zielstra, Hochmair, & Neis (2013)	2006 to 2012	US^	TIGER/Line		x		x						
Barron, Neis, & Zipf (2014)	2007 to 2013	US, Spain, Cameroon	N/A (Framework)	x	x	x	x	x	x	x		x	
Fan, Zipf, Fu, & Neis (2014)	2013	Munich, Germany	ATKIS			x	x	x		x	x		
Forghani & Delavar (2014)	N/A	Tehran, Iran	Municipality of Tehran		x		x	x		x			
Jilani et al. (2014)	N/A	London and East Essex, UK	N/A		x				x		x		
Jokar Arsanjani & Bakillah (2014)	2013	Baden-Württemberg, Germany	N/A (User behavior)	x	x	x	Logistic regression relationship between highly contributed areas and socio-economic variables						
Quattrone, Mashhadi, Quercia, Smith-Clarke, & Capra (2014)	2007 to 2012	London, UK	N/A	x								x	
Tenney (2014)	N/A	Canada^	NRN (2011)		x		x	x	x				
Zhou, Huang, & Jang (2014)	N/A	China	National basic data		x		x	x	x	x			
Ballatore et al. (2015)	2015	Germany and UK	N/A (Framework)		x		Conceptual quality: accuracy, granularity, completeness, consistency, compliance and richness						
Camboim, Meza Bravo, & Sluter (2015)	2015	Brazil	IBGE		x	x	x		x			x	
Dorn, Törnros, & Zipf (2015)	2014	Rhine-Neckar, Germany	ATKIS			x	x		x				
Eckle & De Albuquerque (2015)	N/A	Germany	Map from expert mapper			x	x	x					
Hashemi & Ali Abbaspour (2015)	2014	Wörrstadt, Germany	N/A (Framework)	x	x	x				x			
Hochmair, Zielstra, & Neis (2015)	2013	Portland and Miami, US	Buehler & Pucher (2012)		x		x						
Jokar Arsanjani, Helbich, Bakillah, & Loos (2015)	2007 to 2012	Heidelberg, Germany	N/A	x	x	x						x	
Jokar Arsanjani, Mooney, Zipf, & Schauss (2015)	2013	Germany	GMESUA			x	x		x				
Jokar Arsanjani & Vaz (2015)	2013	European cities	GMESUA			x	x		x				
Kalantari & La (2015)	2013	Victoria, Australia	Victorian governmental data			x	x	x	x	x			

continued on following page

Table 10. Continued

Studies	Time	Study Areas	Reference Data Sources	Data Types			Quality Criteria or Methodology						
				POI	Line	Poly	C	PA	AA	LC	SA	TQ	L
Mashhadi, Quattrone, & Capra (2015)	2007 to 2012	London, UK	Navteq and Yelp	x			x					x	
Meier (2015)	N/A	Waterloo, Canada	NRN		x		x	x					
Mohammadi & Malek (2015)	2012	Tehran, Iran	N/A	x	x	x		x					
Mullen et al. (2015)	2011	Denver, US	ORNL	x			Non-spatial and spatial regression relationships between demographic characteristics and C and PA of OSM						
Parr (2015)	2006 to 2013	US^	US census and governmental data	x	x	x	The Activity-Context-Geography Model						
Sehra, Singh, & Rai (2015)	N/A	India	Ground data by smartphone		x		x	x	x				
Vaz & Jokar Arsanjani (2015)	2013	Toronto, Canada	DMTI Spatial Inc.			x			x				
El-Ashmawy (2016)	N/A	Saudi Arabia	Self-collected surveying data	x	x	x		x					
Yang, Fan, & Jing (2016)	2010 to 2014	Germany, France and UK	N/A (User behavior)	x	x	x	Use practice, skill and motivation as themes to identify the contributors' level of expertise						
Zhao, Zhou, Li, & Xing (2016)	2006 to 2014	Berlin, Germany and Pakistan	N/A (Data trust and vandalism)	x	x	x	Trustworthiness (e.g., contributor reputations)						

Note. ^: a national study; abbreviations:
AA Attribute Accuracy
ATKIS German Authority Topographic-Cartographic Information System
BD TOPO Topographic datasets from the French National Institute of Geography
BKG German Federal Agency for Cartography and Geodesy
C Completeness
CLC Corine Land Cover database
GMESUA Global Monitoring for Environment and Security Urban Atlas
IBGE Brazilian Institute of Geography and Statistics
ITN Integrated Transport Network
L Lineage
LC Logical Consistency
N/A Not Available
NMA National Mapping Agencies (South Africa)
NRN National Road Network (Canada)
ORNL Oak Ridge National Laboratory
OS Ordnance Survey, the national mapping agency for Great Britain
PA Positional Accuracy
SA Semantic Accuracy
TIGER/Line Topologically Integrated Geographic Encoding and Referencing (US)
TQ Temporal Quality
VMD Vector Map District