A Review of Volunteered Geographic Information Quality Assessment Methods

Hansi Senaratne*, Amin Mobasheri, Ahmed Loai Ali, Cristina Capineri, Mordechai (Muki) Haklay

With the ubiquity of advanced web technologies and location-sensing hand held devices, citizens regardless of their knowledge or expertise, are able to produce spatial information. The phenomena is known as Volunteered Geographic Information (VGI). During the last decade VGI has been used as a data source supporting a wide range of services such as environmental monitoring, events reporting, human movement analysis, disaster management etc. However, these volunteer contributed data also come with varying *quality*. Reasons for this are: data is produced by heterogeneous contributors, using various technologies and tools, having different level of details and precision, serving heterogeneous purposes, and a lack of gatekeepers. Crowd-sourcing, social, and geographic approaches have been proposed and later followed to develop appropriate methods to assess the quality measures and indicators of VGI. In this paper, we review various quality measures and indicators for selected types of VGI, and existing quality assessment methods. As an outcome, the paper presents a classification of VGI with current methods utilized to assess the quality of selected types of VGI. Through these findings we introduce *data mining* as an additional approach for quality handling in VGI.

Keywords: Volunteered Geographic Information; Spatial Data Quality; Spatial Data Applications

This is an accepted manuscript of an article published by Taylor & Francis in the International Journal of Geographical Information Science on 31 May 2016, available online: `http://www.tandfonline.com/doi/full/10.1080/13658816.2016.1189556`.

To cite this article:
Hansi Senaratne, Amin Mobasheri, Ahmed Loai Ali, Cristina Capineri & Mordechai (Muki) Haklay (2016): A review of volunteered geographic information quality assessment methods, International Journal of Geographical Information Science, DOI: 10.1080 13658816.2016.1189556

*Corresponding author. Email: hansi.senaratne@uni-konstanz.de

1. Introduction

Volunteered Geographic Information (VGI) is where citizens, often untrained, and regardless of their expertise and background create geographic information on dedicated web platforms (Goodchild 2007), e.g., OpenStreetMap (OSM)[1], Wikimapia[2], Google MyMaps [3], Map Insight [4]and Flickr[5]. In a typology of VGI, the works of Antoniou *et al.* (2010) and Craglia *et al.* (2012) classified VGI based on the type of explicit/implicit geography being captured and the type of explicit/implicit volunteering. In explicit-VGI, contributors are mainly focused on mapping activities. Thus, the contributor explicitly annotates the data with geographic contents (e.g., geometries in OSM, Wikimapia, or Google). Data that is implicitly associated with a geographic location could be any kind of media: text, image, or video referring to or associated with a specific geographic location. For example, geotagged microblogs (e.g., Tweets), geotagged images from Flicker, or Wikipedia articles that refer to geographic locations. Craglia *et al.* (2012) further elaborated that for each type of implicit/explicit geography and volunteering there are potentially different approaches for assessing the *quality*.

Due to the increased potential and use of VGI (as demonstrated in the works of Chunara *et al.* (2012), Sakaki *et al.* (2010), Fuchs *et al.* (2013), MacEachren *et al.* (2011), Liu *et al.* (2008), McDougall (2009), Bulearca and Bulearca (2010), Jacob *et al.* (2009)), it becomes increasingly important to be aware of the quality of VGI, in order to derive accurate information and decisions. Due to a lack of standardization, quality in VGI has shown to vary across heterogeneous data sources (text, image, maps etc.). For example, as seen in Figure 1 a photo of the famous tourist site the Brandenburg Gate in Berlin is incorrectly geotagged in Jakarta, Indonesia on the photo sharing platform Flickr. On the other hand

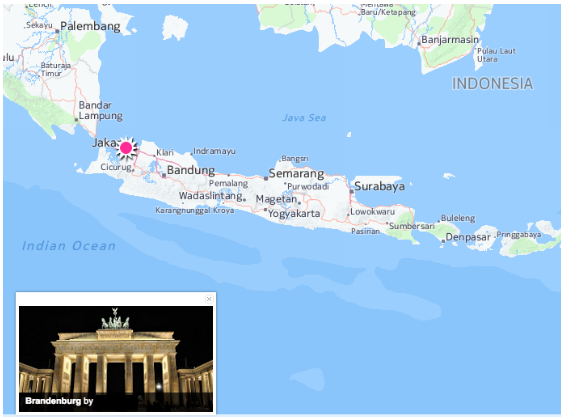

Figure 1.: A photo of the Brandenburg Gate in Berlin is incorrectly geotagged in Jakarta, Indonesia on the popular photo sharing platform Flickr.

OSM has also shown heterogeneity in coverage between different places (Haklay 2010). These trigger a variable quality in VGI. This can be explained by the fact that humans perceive and express geographic regions and spatial relations imprecisely, and in terms of vague concepts (Montello *et al.* 2003). This vagueness in human conceptualization of

[1]http://www.openstreetmap.org
[2]http://www.wikimapia.org
[3]https://www.google.com/maps/mm
[4]http://www.mapsharetool.com/external-iframe/external.jsp
[5]http://www.flickr.com

location is due not only to the fact that geographic entities are continuous in nature, but also due to the quality and limitations of spatial knowledge (Hollenstein and Purves 2010).

Providing reliable services or extraction of useful information require data with a fitness-for-use quality standard. Incorrect (as seen in Figure 1) or malicious geographic annotations could be minimized in place of appropriate quality indicators and measures for these various VGI contributions.

Goodchild and Li (2012) have discussed three approaches for assuring the quality of VGI: crowd-sourcing (the involvement of a group to validate and correct errors that have been made by an individual contributor), social approaches (trusted individuals who have made themselves a good reputation with their contributions to VGI can for example act as gatekeepers to maintain and control the quality of other VGI contributions), and geographic approaches (use of laws and knowledge from geography, such as Tobler's first law to assess the quality). Many works have developed methods to asses the quality of VGI based on these approaches.

In this paper we present an extensive review of the existing methods in the state-of-the-art to assess the quality of map-, image-, and text-based VGI. As an outcome of the review we identify *data mining* as one more stand alone approach to assess VGI quality by utilizing computational processes for discovering patterns and learning purely from data, irrespective of the laws and knowledge from geography, and independent from social or crowd-sourced approaches. Extending the spectrum of approaches will sprout more quality assessment methods in the future, especially for VGI types that have not been extensively researched so far. To the best of our knowledge surveys on existing methods have not been done so far. This review provides an overview of methods that have been built based on theories and discussions in the literature. Furthermore, this survey gives the reader a glimpse to the practical applicability of all identified approaches. The remainder of this paper unfolds as follows: In section 2 we describe the different quality measures and indicators for VGI. In section 3 we describe the main types of VGI that we consider for our survey, and in section 4 we describe the methodology that was followed for the selection of literature for this survey. Section 5 summarizes the findings of the survey, and section 6 discusses the limitations and future research perspectives. Lastly we conclude our findings in section 7.

2. Measures and Indicators for VGI Quality

Quality of VGI can be described by quality *measures* and quality *indicators* (Antoniou and Skopeliti 2015). Quality measures, mainly adhering to the ISO principles and guidelines refer to those elements that can be used to ascertain the discrepancy between the contributed spatial data and the ground truth (e.g., completeness of data) mainly by comparing to authoritative data. When authoritative data is no longer usable for comparisons, and the established measures become no longer adequate to assess the quality of VGI, researchers have explored more intrinsic ways to assess VGI quality by looking into other proxies for quality measures. These are called quality indicators, that rely on various participation biases, contributor expertise or the lack of it, background, etc., that influence the quality of VGI, but cannot be directly measured (Antoniou and Skopeliti 2015). In the following these quality measures and indicators are described in detail. The review of quality assessment methods in section 5 is based on these various quality measures and indicators.

2.1. *Quality Measures for VGI*

ISO[1] (International Standardisation Organisation) defined geographic information quality as *totality of characteristics of a product that bear on its ability to satisfy stated and implied needs*. ISO/TC 211[2] (Technical Committee) developed a set of international standards that define the measures of geographic information quality (standard 19138, as part of the meatadata standard 19115). These quantitative quality measures are: completeness, consistency, positional accuracy, temporal accuracy and thematic accuracy.

Completeness describes the relationship between the represented objects and their conceptualizations. This can be measured as the absence of data (errors of omission) and presence of excess data (errors of commission). Consistency is the coherence in the data structures of the digitized spatial data. The errors resulting from the lack of it are indicated by (i) conceptual consistency, (ii) domain consistency, (iii) format consistency, and (iv) topological consistency. Accuracy refers to the degree of closeness between a measurement of a quantity and the accepted true value of that quantity, and it is in the form of positional accuracy, temporal accuracy and thematic accuracy. Positional accuracy is indicated by (i) absolute or external accuracy, (ii) relative or internal accuracy, (iii) gridded data position accuracy. Thematic accuracy is indicated by (i) classification correctness, (ii) non-quantitative attribute correctness, (iii) quantitative attribute accuracy. In both cases, the discrepancies can be numerically estimated. Temporal accuracy is indicated by (i) accuracy of a time measurement: correctness of the temporal references of an item, (ii) temporal consistency: correctness of ordered events or sequences, (iii) temporal validity: validity of data with regard to time.

2.2. *Quality Indicators for VGI*

As part of the ISO standards, geographic information quality can be further assessed through qualitative quality indicators such as the purpose, usage, and lineage. These indicators are mainly used to express the quality overview for the data. Purpose describes the intended usage of the dataset. Usage describes the application(s) in which the dataset has been utilized. Lineage describes the history of a dataset from collection, acquisition to compilation and derivation to its form at the time of use (Van Oort and Bregt 2005, Hoyle 2001, Guinée 2002). In addition, where ISO standardised measures and indicators are not applicable, we have found in the literature more abstract quality indicators to imply the quality of VGI. These are: trustworthiness, credibility, text content quality, vagueness, local knowledge, experience, recognition, reputation. Trustworthiness is a receiver judgment based on subjective characteristics such as reliability or trust (good ratings on the creations, and the higher frequency of usage of these creations indicate this trustworthiness) (Flanagin and Metzger 2008). In assessing the credibility of VGI, the source of information plays a crucial role, as it is what credibility is primarily based upon. However, this is not straight forward. Due to the non-authoritative nature of VGI, the source maybe unavailable, concealed, or missing (this is avoided by gatekeepers in authoritative data). Credibility was defined by Hovland *et al.* (1953) as the *believability of a source or message, which comprises primarily two dimensions, the trustworthiness (as explained above), and expertise*. Expertise contains objective characteristics such as

[1] `http://www.iso.org/iso/home/standards.htm`
[2] `http://www.isotc211.org/`

accuracy, authority, competence, or source credentials (Flanagin and Metzger 2008). Therefore, in assessing the credibility of data as a quality indicator one needs to consider factors that attribute to the trustworthiness and expertise. Metadata about the origin of VGI can provide a foundation for the source credentials of VGI (Frew 2007). Text content quality (mostly applicable for text-based VGI) describes the quality of text data by the use of text features such as the text length, structure, style, readability, revision history, topical similarity, the use of technical terminology etc. Vagueness is the ambiguity with which the data is captured (e.g., vagueness caused by low resolutions) (De Longueville *et al.* 2010). Local knowledge is the contributors' familiarity to the geographic surroundings that she/he is implicitly or explicitly mapping. Experience is the involvement of a contributor with the VGI platform that she/he contributes to. This can be expressed by the time that the contributor has been registered with the VGI portal, number of GPS tracks contributed (for example in OSM) or the number of features added and edited, or the amount of participation in online forums to discuss the data (Van Exel *et al.* 2010). Recognition is the acknowledgement given to a contributor based on tokens achieved (for example in gamified VGI platforms), and the reviewing of their contributions among their peers (Van Exel *et al.* 2010). Maué (2007) described reputation as a tool to ensure the validity of VGI. Reputation is assessed by, for example the history of past interactions that are happening between collaborators. Resnick *et al.* (2000) described contributors' abilities and dispositions as features where this reputation can be based upon. Maué (2007) further argue that similar to the eBay rating system[1], the created geographic features on various VGI platforms can be rated, tagged, discussed, and annotated, which affects the data contributor's reputation value.

3. Map, Image, and Text based VGI: Definitions and Quality Issues

The effective utilization of VGI is strongly associated with data quality, and this varies depending primarily on the type of VGI, the way data is collected on the different VGI platforms, and the context of usage. The following sections describe the selected forms of VGI: 1) *map*, 2) *image*, and 3)*text*, their uses, and how data quality issues arise. These three types of VGI are chosen based on the methods that are used to capture the data (maps: as gps points and traces, image: as photos, text: as plain text), and because they are the most popular forms of VGI currently used. This section further lays the ground work to understand the subsequent section on various quality measures and indicators, and quality assessment methods used for these three types of VGI.

3.1. *Map-based VGI*

Map-based VGI concerns all VGI sources that include geometries as points, lines and polygons, the basic elements to design a map. Among others, OSM, Wikimapia, Google Map Maker, and Map Insight are examples of map-based VGI projects. However, OSM is the most prominent project due to the following reasons: (i) It aims to develop a free map of the world accessible and obtainable for everyone; (ii) It has millions of registered contributors; (iii) It has active mapper communities in many locations; and (iv) It provides free and flexible contribution mechanisms for data (useful for map provision, routing, planning, geo-visualization, point of interests (POI) search etc.). Thus, during the rest of

[1]`http://ebay.about.com/od/gettingstarted/a/gs_feed.htm`

the article we will discuss OSM as an example for map-based VGI. As in most VGI projects, the spatial dimension of OSM data is annotated in the form of nodes, lines, or polygons with latitude/longitude referencing, and attributes are annotated by tags in the form of key-value pairs. Each tag describes a specific geographic entity from different perspectives. There are no restrictions to the usage of these tags: endless combinations are possible, and the contributors are free to choose the tags they deem appropriate. Nevertheless, OSM provides a set of recommendations of accepted key-value pairs, and if the contributors want their contributions to become a part of the map, they need to follow the agreed-upon standards. This open classification scheme can lead to misclassification and reduction in data quality. Map-based VGI is commonly used for purposes like navigation and POI search. For these purposes the positional accuracy and the topological consistency of the entities are as important as their abstract locations. The other dimension is the attribute accuracy, where the annotations associated with an entity should reflect its characteristics without conflicts (e.g., for road tags, oneway=true and two-way=true). In OSM, the loose contribution mechanisms result in problematic classifications that influence the attribute accuracy. In addition to accuracy, providing reliable services is affected by data completeness; features, attribute, and model completeness. Whether a map includes all the required features, whether a feature is annotated with a complete set of attributes, and if the model is able to answer all possible queries, all these points are related to the completeness quality measure. Especially due to the lack of ground-truth data for comparison, assessing VGI completeness still raises some challenges.

3.2. *Image-based VGI*

Image-based VGI is mostly produced implicitly within portals such as Flickr, Panoramio, Instagram etc., where contributors take pictures of a particular geographic object or surrounding with cameras, smart phones, or any hand held device, and attach a geospatial reference to it. These objects/surroundings can be spatially referenced either by giving geographic coordinates and/or user-assigned geospatial descriptions of these photographs in the form of textual labels. These photo sharing websites have several uses such as environmental monitoring (Fuchs *et al.* 2013), pedestrian navigation (Robinson *et al.* 2012), event and human trajectory analysis (Andrienko *et al.* 2009), for creating geographical gazetteers (Popescu *et al.* 2008), or even to complement institutional data sources in your locality (Milholland and Pultar 2013).

Tagging an image is a means of adding metadata to the content in the form of specific keywords to describe the content (Golder and Huberman 2006), or in the form of geographic coordinates (geotagging) to identify the location linked to the image content (Valli and Hannay 2010). There exist several approaches to geotag an image: record the geographic location with the use of an external GPS device, with an in-built GPS (in many of the modern digital cameras, smart phones), or manually positioning the photo on a map interface.

Not only the GPS precision and accuracy errors resulting from various devices, but also other factors influence the quality of image-based VGI. For example, instead of stating the position from where the photo was taken (photographer position) some contributors tend to geotag the photo with the position of the photo content, which could be several kilometers away from where the photo originated causing positional accuracy issues (as also discussed in Keßler *et al.* (2009)). This is a problem when we want to utilize these photos for example in human trajectory analysis. Furthermore, due to the lack of sufficient spatial knowledge contributors sometimes incorrectly geotag their photographs (Figure

1), also in lower geographic resolutions (in case of Flickr, some contributors do not zoom enough to the street level, instead they zoom up to country or city level to geotag their photos). Or some contributors geotag and textually label random irrelevant photos for actual events, causing the users to doubt the trustworthiness of the content. Such content are not fit for use for tasks such as disaster management, environmental monitoring, or pedestrian navigation. Citizen Science Projects such as GeoTag-X[1] have in place machine learning and crowd-sourcing methods to discover unauthentic material and clean them.

3.3. *Text-based VGI*

Text-based VGI (typically microblogs) is mostly produced implicitly on portals such as Twitter, Reddit or various Blogs, where people contribute geographic information in the form of text by using smart phones, PCs, or any hand held devices. Twitter for example is used as an information foraging source (MacEachren *et al.* 2011), in journalism to disseminate data to the public in near real-time basis (O'Connor 2009, Castillo *et al.* 2011), detect disease spreading (Chunara *et al.* 2012), event detection (Bosch *et al.* 2013), and for gaining insights on social interaction behavior (Huberman *et al.* 2008) or trajectories of people (Andrienko *et al.* 2013, Senaratne *et al.* 2014).

In text-based VGI, the spatial reference can be either in the text, where the contributor refers to a place-name (e.g., 'Lady Gaga is performing in New York today'), or the spatial reference can be the geotag where the tweet is originating from. While some people contribute meaningful information most others use these mediums to express personal opinions, moods, or for malicious aims such as bullying or trolling to harass other users. Gupta and Kumaraguru (2012) conducted a study to investigate how much information is credible and therefore useful, and how much information is spam, on Twitter. They found that 14% of Tweets collected for event analysis were spam, while 30% of the Tweets contained situational awareness information, out of which only 17% of the total tweets contained credible situational awareness information. Such spam makes it difficult to derive useful information that could be of interest for the above named use-cases. Therefore quality analysis of these data is important to filter out the useful information, and disregard the rest. Other than the inherent GPS errors in devices, a bigger role for quality issues is played by the contributor herself/himself based on the information she/he provides. Also due to the lack of spatial knowledge of some contributors the location is incorrectly specified, and at times at a low resolution (in the Twitter interface on PCs the contributor can specify the location not only at the city level, but also at a more coarse state level). Sometimes if the contributor is writing about an event that takes place a few hundred kilometers away from her position, she would geotag her content with the location of the event rather than her position. Or the other way around.

A summary of quality assessment methods for these VGI types is presented in Section 5.

4. The Literature Review Methodology

This review provides an overview of the state-of-the-art methods to assess the quality of selected types of VGI. To achieve this goal we breakdown our review in to three categories. Firstly, we show how the topic of quality assessment within map, image, and text VGI has evolved over the years since the birth of VGI in 2007 until the time of writing this

[1]`http://geotagx.org/`

article (mid of 2015). Secondly, the reviewed papers are classified according to the type of quality measure or indicator that is assessed within each of the papers. Thirdly, all the quality measures and indicators that are addressed within each of the reviewed papers are classified with the different methods utilized to assess them.

We used the following strategy to select the literature for our review. We used Google Scholar to search for papers that include the following terms in their title or abstract: *data quality assessment*, *methods and techniques*, *uncertainty*, *volunteered geographic information*, *map*, *microblog*, *photo*. This query resulted in 425 research papers. We sorted the search results according to the Google Scholar relevance ranking [1]. This relevance ranking follows a combined ranking algorithm that contains a weighting for the full text of each article, author of article, publisher, and how often the article has been cited in other scholarly articles. We refined our collection of papers by filtering out the papers according to the following criteria: 1) papers were published from 2007, 2) papers should describe quality assessment methods, or techniques, or tools, 3) a latest paper was selected when multiple versions of similar methods were available from the same research group. Citizen Science research studies are not considered in this review. As such, we selected 56 papers in total.

Figure 2 shows the distribution of the reviewed papers for VGI quality assessment methods. Evidently, the publication of papers on this topic gained momentum in 2010, for the most part papers discuss methods for map-based VGI.

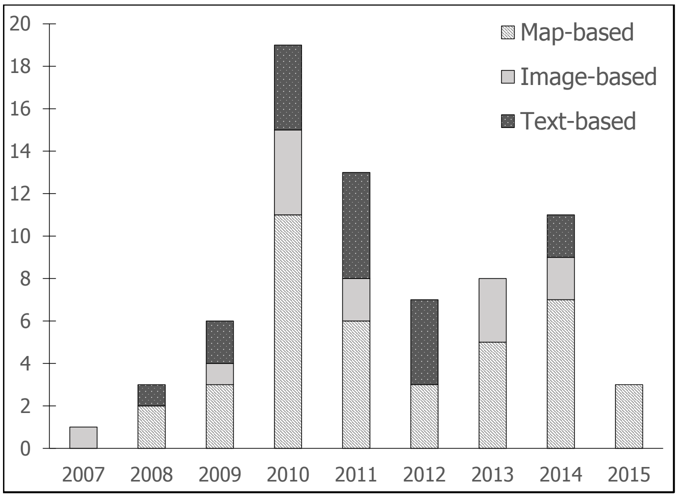

Figure 2.: Distribution of the surveyed papers

5. Existing Methods for Assessing the Quality of VGI

We have reviewed state-of-the-art methods to assess various quality measures and indicators of VGI. Within this review, a method is considered to be a systematic procedure that

[1] https://scholar.google.com/scholar/about.html

is followed to assess the quality measures and quality indicators. For example, comparing with satellite imagery is a method to assess the positional accuracy of maps. The found methods have been mostly conceptually implemented for a particular usecase. These methods have been reviewed mainly based on the type of VGI, the quality measures and indicators supported, and the approaches followed to develop the method.

5.1. *Distribution of Selected Literature*

Out of the 56 papers that we reviewed, 40 papers discuss methods for assessing the quality of map-based VGI, in most cases taking OSM data as the VGI source. 18 papers introduce methods for text-based VGI taking mainly Twitter, Wikipedia, and Yahoo! answers as the VGI source. 13 papers introduce methods for image-based VGI taking Flickr and Panoramio as their VGI source. In reference to Craglia *et al.* (2012)'s typology of VGI with the reviewed papers, most quality assessment work is done on explicit VGI and lesser amount of work is done on implicit VGI, although implicit VGI due to its very nature has more concerns regarding its quality.

5.2. *Type of Quality Measures, Indicators, and their Associated Methods*

We have found 17 quality measures and indicators (7 measures and 10 indicators) that are addressed within the 56 papers we surveyed. In Table 1 we have classified these surveyed papers according to the type of quality measures and indicators, and the type of VGI. We found that papers particularly focusing on map-based VGI are clearly using only ISO standardized measures for quality assessment, whereas text-based VGI have been assessed only on the credibility, text content quality, and vagueness. Image-based VGI have been assessed in several papers on the positional/thematic accuracy, credibility, vagueness, experience, recognition, and reputation. Within these 56 papers we came across 30 methods to assess these quality measures and indicators.

These quality measures/indicators gather previously discussed spatial data quality elements in the literature, but also extend the previous categorizations such as Thomson *et al.* (2005), to include further spatial data quality indicators such as reputation, text content quality, or experience. A classification of the VGI quality measures and indicators according to the type of quality assessment methods and the type of VGI used in the respective applications is presented in Table 2. The sparse cells in the table indicate the quality measures/indicators that have not been explored excessively. We have further classified these methods according to the approach categorization by Goodchild and Li (2012). In addition to their categorization, we have also found methods based on the data mining approach.

	Quality measures and indicators																
Papers	Positional accuracy	Thematic accuracy	Topological consistency	Completeness	Temporal accuracy	Geometric accuracy	Semantic accuracy	Lineage	Usage	Credibility	Trustworthiness	Content quality	Vagueness	Local knowledge	Experience	Recognition	Reputation

Jacobs et al.(2007)	$\bullet$																
Agichtein et al.(2008)												$\diamond$					
Schmitz et al.(2008)			$\star$														
Mummidi&Krumm (2008)		$\star$															
Hasan et al.(2009)												$\diamond$					
Kounadi (2009)	$\star$																
Ather (2009)	$\star$			$\star$													
De Longueville et al.(2010)													$\bowtie$				
Bishr&Janowicz(2010)											$\bowtie$						
Mendoza et al.(2010)										$\diamond$							
Haklay(2010)	$\star$			$\star$													
Ciepluch(2010)	$\star$			$\star$													
Corcoran et al.(2010)			$\star$														
Girres&Touya (2010)		$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$								
Haklay et al. (2010)	$\star$																
Poser&Dransch (2010)		$\bowtie$															
Brando&Bucher (2010)	$\bowtie$	$\bowtie$	$\bowtie$	$\bowtie$	$\bowtie$												
Huang et al. (2010)											$\bowtie$						
De Tré et al. (2010)	$\star$	$\star$															
Al Bakri&Fairbairn (2010)	$\star$																
van Exel et al. (2010)															$\bowtie$	$\bowtie$	$\bowtie$
Ciepluch et al. (2011)			$\star$														
Neis et al. (2011)			$\star$														
Codescu(2011)		$\star$															
Castillo et al.(2011)										$\diamond$							
Becker et al (2011)												$\diamond$					
Canini et al.(2011)										$\diamond$							
Ostermann&Spinsanti (2011)										$\bowtie$							
Kessler et al. (2011)								$\star$									
O'Donovan et al.(2012)										$\diamond$							
Kang et al.(2012)										$\diamond$							
Gupta et al.(2012)										$\diamond$							
Morris et al. (2012)										$\diamond$							
Helbich et al.(2012)	$\star$																
Mooney&Corcoran(2012)		$\star$															
Koukoletsos et al. (2012)				$\star$													
Kessler&deGroot(2013)		$\star$	$\star$	$\star$													
Senaratne et al.(2013)	$\bullet$									$\bullet$							
Zielstra&Hochmair(2013)	$\bullet$																
Canavosio-Zuzelski et al.(2013)	$\star$																
Hecht et al.(2013)				$\star$													
Vandecasteele&Devillers(2013)		$\star$															
Jackson et al. (2013)	$\star$			$\star$													
Foody et al. (2014)		$\bullet$															
Barron et al.(2014)			$\star$	$\star$													
Siebritz(2014)			$\star$														

Wang et al.(2014)			⋆														
Fan et al.(2014)	⋆			⋆													
Tenney(2014)	⋆			⋆													
Ali et al.(2014)		⋆															
Bordogna et al. (2014)													●◇		●	●	●
Forghani &Delavarl(2014)			⋈														
Hollenstein&Purves(2014)	●																
Arsanjani (2015)		⋆															
Vandecasteele&Devillers (2015)							⋆										
Hashemi&Abbaspour (2015)			⋆														

Table 1.: Classification of the reviewed papers according to the quality measures and indicators. ⋆ = map-based, ● = image-based, and ◇ = text-based VGI. While ⋈ = all types of VGI.

	Type of approaches and methods																													
	Geographic												Social							Crowdsourcing	Data mining									
	Compare with reference data	Line of sight	Formal specifications	Semantic consistency check	Geometrical analysis	Intrinsic data check	Integrity constraints	Automatic tag recommendation	Geographic proximity	Time between observations	Automatic scale capturing	Geographic familiarity	Manual inspection	Manual inspection/annotation	Manual annotation	Comparing limitation with previous evaluation	Linguistic decision making	Meta-data analysis	Tokens achieved, peer reviewing	Applying Linus law	Possibilistic truth value	Cluster analysis	Latent class analysis	Correlation statistics	Automatic detection of outliers	Regression analysis	Supervised classification	Feature classification	Provenance vocabulary	Heuristic metrics/fuzzy logic
Positional accuracy	⋆ ● ◇	●	⋆ ● ◇										●	●						⋆	⋆									
Thematic accuracy	⋆ ● ◇		⋆ ● ◇	⋆											⋆							⋆	●						⋆	
Topological consistency	⋆ ● ◇		⋆ ● ◇	⋆	⋆	⋆	⋆																						⋆	⋆ ● ◇
Complet-ness	⋆ ● ◇		⋆ ● ◇			⋆		⋆																					⋆	
Temporal accuracy																								⋆						

Geometric accuracy	⋆																													
Semantic accuracy	⋆							⋆																						
Lineage													⋆																⋆	
Usage																⋆														
Credibility	⋆ ● ◇	●																									◇	◇		
Trust									⋆ ● ◇	⋆ ● ◇															⋆ ● ◇					
Content quality																	● ◇									◇		◇		
Vagueness											⋆ ● ◇				⋆ ● ◇															
Local knowledge												⋆ ● ◇																		
Experience																	● ◇	⋆ ● ◇												
Recognition																	● ◇		⋆ ● ◇											
Reputation																	● ◇													

Table 2.: Quality measures and indicators are classified according to the type of methods to assess them, and the types of VGI. Methods are further classified according to the quality assessment approaches. ⋆ = map-based, ● = image-based, and ◇ = text-based VGI.

5.2.1. Quality Assessment in Map-based VGI

Positional Accuracy

In the works of Kounadi (2009), Ather (2009), Haklay (2010), Ciepłuch *et al.* (2010), Al-Bakri and Fairbairn (2010), Zandbergen *et al.* (2011), Helbich *et al.* (2012), Jackson *et al.* (2013), Fan *et al.* (2014), Tenney (2014), Brando and Bucher (2010), Al-Bakri and Fairbairn (2010), authors employ officially gathered reference datasets to assess the positional accuracy of map-based VGI (mostly OSM data) by comparison. The comparison with reference data method has been further employed for the assessment of thematic accuracy (Girres and Touya 2010, Poser and Dransch 2010, Kounadi 2009, Brando and Bucher 2010, Arsanjani *et al.* 2015), completeness (Haklay 2010, Ciepłuch *et al.* 2010, Kounadi 2009, Ather 2009, Ciepłuch *et al.* 2011, Hecht *et al.* 2013, Jackson *et al.* 2013, Fan *et al.* 2014, Tenney 2014, Brando and Bucher 2010), geometric accuracy (Girres and Touya 2010). For geometric accuracy OSM objects of same structure were manually matched. This manual approach was preferred over an automated approach to avoid any processing errors.

Haklay *et al.* (2010) applied the Linus Law and found out that higher the number of contributors on a given spatial unit on OSM, higher the quality. This study shows that comparison to reference datasets isn't the only way to assess the quality of OSM data as done in many use-cases.

De Tré *et al.* (2010) uses a Possibilistic Truth Value (PTV) as a normalized possibility distribution to determine the uncertainty of the POIs being co-located. The uncertainty regarding the positioning of a POI is primarily caused by the imprecision with which the POI are positioned on the map interface. The proposed technique further semantically checks and compares the closely located POIs. Their method helps to identify redundant VGI, and fuse the redundancies together. Furthermore, this approach has been applied to also assess the thematic accuracy of map-based VGI.

In a rather different approach, Canavosio-Zuzelski *et al.* (2013) perform a photogrammetric approach for assessing the positional accuracy of OSM road features using stereo imagery and a vector adjustment model. Their method applies analytical measurement principles to compute accurate real world geo-locations of OSM road vectors. The proposed approach was tested on several urban gridded city streets from the OSM database with the results showing that the post adjusted shape points improved positional accuracy by 86%. Furthermore, the vector adjustment was able to recover 95% of the actual positional displacement present in the database.

Brando and Bucher (2010) present a generic framework to manage the quality of ISO standardized quality measures by using formal specifications and reference datasets. Formal specifications facilitate the assurance of quality in three manners with means of integrity constraints: i) support on-the-fly consistency checking, ii) comparison to external reference data, iii) reconcile concurrent editions of data. However, due to a lack of proof of concept the practical applicability of this approach is difficult to conceive.

Topological Consistency

The topological consistency in OSM data is assessed mainly using intrinsic data checks to detect and alleviate problems occurring through for example overlapping features or overshoots and undershoots in the data (also known as dangles where start and end point of two different lines should meet but do not, due to bad practices in digitization). The authors Schmitz *et al.* (2008), Neis *et al.* (2011), Barron *et al.* (2014), Siebritz (2014) have demonstrated that for each of these measures a separate topology integrity rule can be designed and applied. Further, based on the definition of planar and non-planar topological properties Corcoran *et al.* (2010) and Da Silva and Wu (2007) have used geometrical analysis methods to assess the topological consistency of the OSM data. In another work, the concept of spatial similarity in multi-representations have been employed in order to perform both extrinsic and intrinsic quality analysis (Hashemi and Abbaspour 2015). The authors discuss that their method could be efficiently applied to VGI data for the purpose of vandalism detection. Other studies have also focused on evaluating the topological consistency of OSM data with a focus on road network infrastructures (Will 2014). In Wang *et al.* (2014) and Girres and Touya (2010) the authors have used the Dimensional Extended nine-Intersection Model (DE-9IM) in order to compute the qualitative spatial relation between road objects in OSM. This method and model allows them to check for topological inconsistencies and be able to locate the junctions of roads in order to, for example generate expected road signs.

Thematic Accuracy and Semantic Accuracy

Mooney and Corcoran (2012) points out that most errors in OSM are caused by manual annotation by contributors who sometimes misspell the feature values. Addressing this issue, Codescu *et al.* (2011), Vandecasteele and Devillers (2013), Ali *et al.* (2014) have developed semantic similarity matching methods, which automatically assess the contributor annotation of features in OSM according to the semantic meaning of such features. In the work of Girres and Touya (2010), they found semantic errors were mainly due to the mis-specification of roads. For example: roads that were classified as 'secondary' in the reference dataset were classified as 'residential', or 'tertiary' by contributors in OSM data. The reasons for these inaccuracies as seen by authors are the lack of a standardized classification, looseness for contributors to enter tags and values that are not present in the OSM specification, lack of naming regulations w.r.t. for example capitalization or prefixes. The authors emphasize the need for standardized specifications to improve semantic and attribute accuracy of OSM data.

Furthermore, in regard to semantic accuracy of map-based VGI, Vandecasteele and Devillers (2015) introduced a tag recommender system for OSM data which aims to improve the semantic quality of tags. OSMantic is a plugin for the Java OpenStreetMap editor which automatically suggests relevant tags to contributors during the editing process. Mummidi and Krumm (2008) use clustering methods on Microsoft's Live Search Maps[1] to group user contributed pushpins of POIs that are annotated with text. Frequent text phrases that appear in one cluster but infrequently in other clusters help to increase the confidence that the particular text phrase describes a POI.

Completeness

Koukoletsos *et al.* (2012) propose to use a feature-based automated matching method for linear data using reference datasets. Barron *et al.* (2014) and Girres and Touya (2010) use intrinsic data checks to record the statistics of the number of objects, attributes, and values, thereby keeping track of all omissions and commissions to the database.

Temporal Accuracy

Very few works exist to assess the temporal accuracy. We reviewed the works of Girres and Touya (2010) where they use statistics to observe the correlations of the number of contributors to the mean capture date, and to the mean version of the capture object in order to assess how many objects are updated. Their results show a linear increase of the mean date, and the mean version of captured object in relation to the number of contributors in the chosen geographic area. Concluding results show higher the number of contributors, more recent the objects were, and the more up-to-date the objects were.

Lineage, Usage, Purpose

In Keßler *et al.* (2011), following a data oriented approach with a focus on the origins of specific data items, their provenance vocabulary explicitly shows the lineage of data features of any online data. They base their provenance approach on Hartig (2009) on 'provenance information in the web of data'. Their approach allows them to classify OSM features according to recurring editing and co-editing patterns. To keep track of the data lineage, Girres and Touya (2010) urge the need for moderators who have control over screening the contributions (as in Wikipedia) for necessary source information. They further analyze the usage of data by comparing the limitations that were observed in previous evaluations of map-based VGI.

As a generic approach to assess ISO standardized quality indicators, (Keßler and de Groot 2013) propose Trust as a proxy to measure the topological consistency, thematic accuracy, and completeness in these map data based on data provenance, a method which

[1] `http://maps.live.com`

relies on trust indicators as opposed to ground truth data.

5.2.2. Quality Assessment in Image-based VGI

Positional Accuracy and Credibility

Jacobs *et al.* (2007) explored the varying positional accuracy of photos by matching photos with ancillary satellite imagery. They localize cameras based on satellite imagery that correlates with the camera images taken at a known time. Their approach helps where it is important to know the accurate location of the photographer instead of the target object. Zielstra and Hochmair (2013) on the other hand compared the geotagged positions of photos to the manually corrected camera position based on the image content. Their results indicate better positional accuracy for Panoramio photos compared to Flickr photos. Hollenstein and Purves (2010) assessed the positional accuracy of such photos by manually inspecting these photos for their correspondence between the tagged geographic label and geotagged position. Senaratne *et al.* (2013) assessed the positional accuracy of Flickr photos by computing a line of sight between the camera position and the target position based on in-between surface elevation data. They further manually inspected the geographic label against the geographic location. The results are used as a reference of quality for contributor and photo features of Flickr, and thereby used to derive credibility indicators.

Thematic Accuracy

Foody *et al.* (2014) use Geowiki as the data source, where it contains a series of satellite imagery. Volunteered contributors were given the task to label the land use categories in these satellite imagery from a pre-defined set of labels. The accuracy of the labeling was assessed through conducting a latent class analysis (LCA). LCA allows the analyst to derive an accuracy measurement of the classification when there are no reference datasets available to compare with. The authors further emphasize that this method can be applied to image-based VGI. Further, their approach characterizes the volunteers based on the accuracy of their labels of land use classes. This helps to ultimately determine the volunteer quality.

On a related work, Zhang and Kosecka (2006) used feature-based geometric matching using the image recognition software SIFT (Lindeberg 2012) to localize sample photos in urban environments. Although their work was not based on VGI, this is a potential approach to solve quality related issues within image-based VGI.

5.2.3. Quality Assessment in Text-based VGI

Quality of text-based VGI has been mainly assessed through the credibility of such data based on contributor, text, and content features, and through the text content quality.

Credibility

Relating to a social approach of quality analysis, Mendoza *et al.* (2010) found out that rumors on Twitter tend to be more questioned by the Twitter community during an emergency situation. They further indicate that the Twitter community acts as a collaborative filter of information.

Castillo *et al.* (2011) employed users on mechanical turk[1] to classify pre-classified 'news-worthy events' and 'informal discussions' on Twitter according to several classes

[1] https://www.mturk.com

of credibility (i. almost certainly true, ii. likely to be false, ..). This is then used in a supervised classification to evaluate which Tweets belong to these different classes of credibility. This helped the authors to derive credibility indicators. The user features such as average status count or the number of followers among others were found to be the top ranked user-based credibility features.

The work of Gupta and Kumaraguru (2012) is similar to Castillo *et al.* (2011), and follows a supervised feature classification PageRank like method to propagate the credibility on a network of Twitter events. They use event graph-based optimization to enhance the trust analysis at each iteration that updates the credibility scores. A credible entity (node) links with a higher weight to more credible entities than to non-credible ones. Their approach is similar to that of Castillo *et al.* (2011), but the authors proposed a new technique to re-rank the Tweets based on a Pseudo Relevance Feedback.

Canini *et al.* (2011) divided credibility into implicit and explicit credibility. Implicit credibility is the perceived credibility of Twitter contributors, and is assessed by Twitter users by evaluating an external data source together with the Tweeters content topicality and its relevance to the context, and social status (follower/ status counts). Explicit credibility is evaluated by ranking Tweeters (Twitter contributors) on a scale from 1 to 5 based on their trustworthiness. End result is a ranking recommendation system on whom to follow on Twitter regarding a particular topic.

O'Donovan *et al.* (2012) provided an analysis of the distribution of credibility features in four different contexts in the Twitter network: diversity of topics, credibility, chain length and dyadic pairs. The results of their analysis indicate that the usefulness of credibility features depends on the context in question. Thus the presence of a credibility feature alone is not good enough to evaluate the credibility of the context, but rather a particular combination of different credibility features that are 'suitable' for the context in question.

Morris *et al.* (2012) designed a pilot study with participants (with no technical background) to extract a list of features that are useful to make their credibility judgments. Finally to run the survey, the authors sent the survey to a sample of Twitter users in which they were asked to assess how each feature impacts their credibility judgment on a five-point scale. Their findings indicate that features such as verified author expertise, re-tweets from someone you trust, or author is someone you follow have higher credibility impact. These features differ somewhat to the features extracted through the supervised classification of Castillo *et al.* (2011). These features were further ranked according to the amount of attention received by Twitter users.

Kang *et al.* (2012) defined three different credibility prediction models and studied how each model performs in terms of credibility classification of Twitter messages. These are: 1. social model, 2. content-based model, and 3. hybrid model (based on different combinations of the two previous models). The social model relies on a weighted combination of credibility indicators from the underlying social network (e.g., re-tweets, no. of followers). The content-based model identifies patterns and tweet properties that lead to positive reactions such as re-tweeting or positive user ratings, by using a probabilistic language-based approach. Most of these content-based features are taken from Castillo *et al.* (2011). The main results from the paper indicate that the social model outperformed all other models in terms of predication accuracy, and that including more features in the predication task doesn't mean a better predication accuracy.

Text Content Quality

Agichtein *et al.* (2008) describe a generic method for all text-based social media data.

They use three inputs for a feature classifier to determine the content quality: 1. textual features (e.g., word n-grams up to length 5 that appears in the text more than 3 times, semantic features such as punctuations, typos, readability measures, avg. no. of syllables per word, entropy of word lengths, grammarticality), 2. user relationships (between users and items, user intuition such as good answers are given by good answerers, and vote for other good answerers), 3. usage statistics (no. of clicks on an item, dwell time on content).

Becker *et al.* (2011) use a two tier approach for the quality analysis of text-based Twitter data in an event analysis context. To identify the events, they first cluster tweets using an online clustering framework. Subsequently, they use three centrality-based approaches to identify messages in the clusters that have high textual quality, strong relevance, and are useful. These approaches are: 1. centroid similarity approach that calculates the cosine similarity of the tf-idf statistic of words, 2. degree centrality method which represents each cluster message as a node in a graph, and two nodes are connected with an edge when their cosine similarity exceeds a predetermined threshold, 3. LexRank approach distributes the centrality value of nodes to its neighbors, and top messages in a cluster are chosen according to their LexRank value.

Hasan Dalip *et al.* (2009) on the other hand use text length, structure, style readability, revision history, and social network as indicators of text content quality in Wikipedia articles. They further use regression analysis to combine various such weighed quality values into a single quality value, that represents an overall aggregated quality metric for text content quality.

Bordogna *et al.* (2014) measure the validity of text data by measuring the number of words, proportion of correctly spelled words, language intelligibility, diffusion of words, and the presence of technical terms as indicators of text content quality. They further explored quality indicators such as experience, recognition and reputation to determine the quality of VGI.

5.2.4. Generic Approaches

As a generic method for all VGI Forghani and Delavar (2014) propose a new quality metric for the assessment of topological consistency by employing heuristic metrics such as minimum bounding geometry area and directional distribution (Standard Deviational Ellipse). Van Exel *et al.* (2010) propose to use contributor related quality indicators such as local knowledge (e.g., spatial familiarity), experience (e.g., amount of contributions), and recognition (e.g., tokens achieved). A conceptual workflow for automatically assessing the quality of VGI in crisis management scenarios was proposed by Ostermann and Spinsanti (2011). VGI is cross-referenced with other VGI types, and institutional ancillary data that are spatially and temporally close. However, in a realistic implementation this combination of different VGI data types for cross referencing is a challenging task due to their heterogeneity. Bishr and Janowicz (2010) propose to use trust together with reputation as a proxy measure for VGI quality, and established the spatial and temporal dimensions of trust. They assert that shorter geographic proximity of VGI observations provide more accurate information as opposed to higher geographic proximity VGI observations (implying that *locals know better*, *the proximate spectator sees more*). On a temporal perspective of trust, they further claim that trust in some VGI develop and decay over time, and that the observation time of an event has an affect on the trust we endow in one's observation. Furthermore, to assess the trust of VGI Huang *et al.* (2010) developed a method to detect outliers in the contributed data. De Longueville *et al.* (2010) proposed two methods to assess the vagueness in VGI. 1. contributor encodes the vagueness of their contributed spatial data in a 0 - 5 scale (e.g., 5 = it's exactly there, 0 =

I don't know where it is. 2. the second type is system created vagueness that is assessed through automatically capturing the scale at which VGI is produced. VGI produced in lower scales is classified as more vague.

Table 2 shows a summary matrix of all quality measures and indicators observed in the literature review, with various methods that can be applied to assess these quality measures/indicators. Following this matrix we can learn which methods can be applied to solve various quality issues within map, text and image-based VGI. However, this should be followed with caution, as we present here only what we discovered through the literature review, and the presented methods could be applied beyond our discovery, and therefore need to be further explored.

6. Discussion and Future Research Perspectives in VGI Quality

VGI is available with tremendous amounts through various platforms, and it is crucial to have methods to ensure the quality of these VGI. The vast amount of data and the heterogeneous characteristics of utilization make the traditional comparison with reference datasets no longer viable in every application scenario (also due to the lack of access to reference data). Based on such characteristics, Goodchild and Li (2012) propose three approaches to ensure the quality of VGI: 1. crowd-sourced, 2. social, and 3. geographic. As seen in Table 2, 20 of the methods we have discovered in the literature fall in to geographic, social, or crowd-sourced approaches. Furthermore, 10 of the methods we discovered fall in to an additional approach: 4. data mining, that helps to assess VGI quality by discovering patterns and learning purely from the data. Data mining can be used as a stand-alone approach, completely independent of the laws and knowledge of geography, and independent from social or crowd-sourced approaches to assess the quality of VGI. For example, the possibilistic truth value method is used to assess the positional uncertainty of POIs based only on the possibility distribution. Similarly, outlier detection, cluster analysis, regression analysis, or correlation statistics methods can be used to assess the data quality by purely discovering and learning data patterns, irrespective of the laws and knowledge from geography. The supervised learning, and feature classification methods that are used to assess the quality of text-based VGI use text, message, and user features to train the classifier. These two machine learning methods we found in the literature once again work irrespective of the laws and knowledge from geography. Therefore, we believe these methods deserve to be represented under an additional approach to assess VGI quality.

We have classified the found methods according to these 4 approaches based on the description of the methods in the literature. By this discovery, we aim to extend Goodchild and Li (2012)'s classification in this survey.

While most methods have been utilized to assess the positional accuracy, thematic accuracy, and topological consistency, fewer methods tackle the rest of the quality measures and indicators we review such as the completeness, temporal accuracy or vagueness. Future work should focus also on other potential approaches to handle quality measures and indicators. Different VGI platforms should clearly communicate to the contributors and the consumers, as to what kind of data that one could contribute. The more precise this is, the more comprehensive it is to the contributor on what is expected in terms of data. As also stated by Antoniou *et al.* (2010), explicit VGI gives a loosely coupled specification(s) of what volunteers can contribute. If these specifications are more rigid the future of VGI can expect higher quality information, although it may be a compromise with lesser contributions. This may further vary depending on the task at hand.

Lower population density positively correlates with fewer number of contributions, thus affecting data completeness or positional accuracy (Neis *et al.* 2013, Haklay 2010, Girres and Touya 2010, Mullen *et al.* 2015). However, more research needs to be done regarding this issue. Hence, a step further in this direction is to derive the socio-economic impacts on OSM data quality. As presented in section 5.2., there have been a number of studies and empirical research performed on the subject of OSM quality. Nevertheless, a solid framework for assessing OSM data is far from being established, let alone a framework of quality measurement for specific application domains. The limitation is that existing measures and indicators (described by ISO) are not inclusive enough to evaluate OSM data. This is mainly because the nature of OSM (and VGI in general) is fundamentally different to what geospatial experts have dealt with so far. Therefore, we argue that there are still research gaps when defining quality measures/indicators and proposing methods to calculate these measures/indicators. In addition, only few studies have been conducted to explore and analyze the differences in quality requirements for different application domains. Therefore, as a recommendation for future research in this topic, we suggest to develop a systematic framework that provides methods and measures to evaluate the fitness for purpose of each VGI type. This would need to not only focus on the analysis of data itself, but also explore the social factors which are the driving forces behind public contributions, and thus considerably affect the quality. For example, one could define a mathematical model based on OSM intrinsic data indicators (e.g., number of contributors, number of edits, etc.) to estimate the quality (e.g., completeness) of data without having reference data at hand. This would enrich and complete the new paradigm of intrinsic quality evaluation, which by far has received less attention from the research community, compared to the common extrinsic quality evaluation: i.e: comparison with reference data.

The utilization of text and image-based VGI still mostly depends on the geotagged content. However, the sparse geotagged content of these two VGI types in most cases represent only a minority of the data. Therefore, generalization based on VGI is still limited and need further demographic studies.

Gamification has become a popular way to involve people to contribute spatial data (Geograph, Foursquare [1], Ingress [2] are some examples). Such gamification approaches have increased participation as well as spatial coverage (Antoniou and Schlieder 2014, Antoniou *et al.* 2010). Due to the clear incentives of this data collection approach (going high up in rankings, collecting badges etc.) this popular method can be used to control the process of collecting more accurate data by incorporating data quality concepts (Yanenko and Schlieder 2014). One way to do that would be to give a ranking to the contributor based on the quality of their collected data. Revealing such rankings of their peers would further encourage the contributors to pay more attention to the quality of their data (peer pressure).

As encouragement mechanisms are required to motivate people to contribute, we should also research methods to make contributors aware of the importance of quality, and secondly to involve the contributors and consumers to maintain the quality of the VGI contents. This can be achieved for example by collaboratively doing quality checks on the data. Such collaborative efforts are presently actively done in OSM, but rather inadvertently done on Flickr or Twitter. As evident from the review, image and text-based VGI have been given far less attention to its quality as compared to map-based VGI. We

[1] https://foursquare.com/
[2] https://www.ingress.com/

see this as mainly due to the complexity of the image and text data types. Comments and discussions associated with image and text contents might be one way to ensure the contribution while systematic analysis of these resources is not a trivial process. Our understanding is that quality assurance methods for text and image-based VGI are still on the phase of experimentation, and therefore need more attention in order to standardize these methods in to practice. This is crucial because more and more text and image-based VGI are being utilized in various applications. Furthermore, the works of Sacha *et al.* (2016), where they introduce a framework that integrates trust and other various quality indicators in a knowledge generation process within the visual analytics paradigm can be adapted in future research to assess and visually analyze quality of VGI. Their framework allows the user to comprehend the associated quality at each step of knowledge generation, and also express their confidence in the findings and insights gained by externalizing their thoughts. This facilitates the user to comprehend the provided quality of data as well as the perceived quality.

As further evident from this review, there is no holy grail that could solve all types of quality issues in VGI. We should be aware of the heterogeneity of these data, and be informed of the existing state-of-the-art to resolve many of the quality issues of VGI, and their limitations. Addressing these limitations and thereby improving the existing methods already paves for new contributions on this topic that should be recognized as valid scientific contributions in the VGI community.

7. Conclusions

In this review of VGI quality, we have taken a critical look at the quality issues within map, image, and text VGI types. The heterogeneity of these VGI types give rise to varying quality issues that need to be dealt with varying quality measures and indicators, and varying methods. As a result of this review, we have summarized the literature in to a list of 30 methods that can be used to assess one or more of the 17 quality measures and indicators that we have come across in the literature for map, image, and text-based VGI respectively. This review further shows the following: 1) a majority of reviewed papers focus on assessing map-based VGI. 2) Though implicit VGI (e.g., text-based Twitter or image-based Flickr) has higher quality concerns in comparison to explicit VGI (e.g., map-based OSM), such explicit VGI has received significantly higher attention to resolve quality issues, compared to implicit VGI. The review shows the increasing utilization of implicit VGI for geospatial research. Therefore, more efforts should be in place to resolve quality issues within these implicit VGI. 3) Mostly ISO standardized quality measures have been used to assess the quality of map-based VGI (OSM). Text-based VGI have been assessed on the credibility, vagueness, and the content quality. Image-based VGI have been assessed on the positional/thematic accuracy, credibility, vagueness, experience, recognition, and reputation. A logical explanation for this is that ISO standardized measures are most often assessed through comparative analysis with ground truth data. For the explicit VGI (e.g., OSM) we can easily realize which ground truth data to look for. However for implicit VGI, it is not straight forward to realize which ground truth data to look for, therefore comparative analysis is not always possible (e.g., topological consistency, or thematic accuracy cannot be directly assessed, as we need to derive the topology or the thematic attributes from the VGI in an additional data processing step). These implicit VGI are further enriched with contributor sentiments and contextual information. Therefore ISO standardized measures alone are not enough to assess the quality of implicit

VGI. This explains the use of indicators such as reputation, trust, credibility, vagueness, experience, recognition, or local knowledge as quality indicators. A lack of standardization of these more abstract quality indicators is a reason why fewer works exist for image and text-based VGI. In addition, the implicit nature of the geography that is contributed in most of these VGI is yet another reason for the insufficiency of quality assessment methods for text and image-based VGI. 4) we have classified the quality assessment methods according to the crowd-sourced, geographic, and social approaches as introduced by Goodchild and Li (2012). We have further discovered data mining as an additional approach in the literature that extends Goodchild and Li (2012)'s classification.

8. Acknowledgments

This work has been partly funded by the SPP programme under grant agreement no. 1335 (ViAMoD), the European Unions's Seventh Framework Programme under grant agreement no. 612096 (CAP4Access), and the German Academic Exchange Service (DAAD). We thank particularly Tobias Schreck, Alexander Zipf, Mohamed Bakillah, and Hongchao Fan for their valuable discussions on the topic.

References

Agichtein, E., *et al.*, 2008. Finding high-quality content in social media. *In*: *Proceedings of the 2008 International Conference on Web Search and Data Mining, 11-12 February*, Palo Alto, California, USA ACM, New York, NY, USA, 183–194.

Al-Bakri, M. and Fairbairn, D., 2010. Assessing the accuracy of crowdsourced data and its integration with official spatial datasets. 317–320.

Ali, A.L., *et al.*, 2014. Ambiguity and Plausibility: Managing Classification Quality in Volunteered Geographic Information. *In*: *Proceedings of the 22nd International Conference on Geographic Information Systems 4-7th November 2014* ACM, New York, NY, USA.

Andrienko, G., *et al.*, 2009. Analysis of community-contributed space-and time-referenced data (example of flickr and panoramio photos). *In*: *IEEE Symposium on Visual Analytics Science and Technology (VAST) 12-13th October 2009*, Atlantic City, NJ, USA IEEE, New Jersey, USA, 213–214.

Andrienko, G., *et al.*, 2013. Thematic patterns in georeferenced tweets through space-time visual analytics. *Computing in Science and Engineering*, 15 (3), 72–82.

Antoniou, V. and Skopeliti, A., 2015. Measures and Indicators of Vgi Quality: AN Overview. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1, 345–351.

Antoniou, V., Morley, J., and Haklay, M., 2010. Web 2.0 geotagged photos: Assessing the spatial dimension of the phenomenon. *Geomatica*, 64 (1), 99–110.

Antoniou, V. and Schlieder, C., 2014. Participation Patterns, VGI and Gamification. 3–6.

Arsanjani, J.J., *et al.*, 2015. Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets. *In*: P.M.M.H. J. Arsanjani A. Zipf, ed. *OpenStreetMap in GIScience*. Springer Switzerland, 37–58.

Ather, A., 2009. A quality analysis of openstreetmap data. *ME Thesis, University College London, London, UK.*

Barron, C., Neis, P., and Zipf, A., 2014. A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, 18 (6).

Becker, H., Naaman, M., and Gravano, L., 2011. Selecting Quality Twitter Content for Events.. *ICWSM*, 11.

Bishr, M. and Janowicz, K., 2010. Can we trust information?-the case of Volunteered Geographic Information. *In*: P.M.C.K. A. Devaraju A. Llaves, ed. *Proceedings of Towards Digital Earth Search Discover and Share Geospatial Data Workshop at Future Internet Symposium, 20th September, 2010*, Vol. 640, Berlin, Germany CEUR-WS.org.

Bordogna, G., *et al.*, 2014. A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science. *Information Sciences*, 258, 312–327.

Bosch, H., *et al.*, 2013. Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering. *IEEE Transactions on Visualization and Computer Graphics*, 19 (12), 2022–2031.

Brando, C. and Bucher, B., 2010. Quality in user generated spatial content: A matter of specifications. *In*: H.P. M. Painho M.Y. Santos, ed. *Proceedings of the 13th AGILE international conference on geographic information science, 11-14th May, 2010*, Guimar aes, Portugal Springer Verlag, 11–14.

Bulearca, M. and Bulearca, S., 2010. Twitter: a viable marketing tool for SMEs. *Global Business and Management Research: An International Journal*, 2 (4), 296–309.

Canavosio-Zuzelski, R., Agouris, P., and Doucette, P., 2013. A photogrammetric approach for assessing positional accuracy of OpenStreetMap© roads. *ISPRS International Journal of Geo-Information*, 2 (2), 276–301.

Canini, K.R., Suh, B., and Pirolli, P.L., 2011. Finding credible information sources in social networks based on content and social structure. *In*: *Proceedings of Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 9-11 October, 2011*, Boston, Massachusetts, USA IEEE, NJ, USA, 1–8.

Castillo, C., Mendoza, M., and Poblete, B., 2011. Information credibility on twitter. *In*: *Proceedings of the 20th international conference on World wide web*, 675–684.

Chunara, R., Andrews, J.R., and Brownstein, J.S., 2012. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *The American Journal of Tropical Medicine and Hygiene*, 86 (1), 39–45.

Ciepłuch, B., *et al.*, 2010. Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. *In*: P.F. N.J. Tate, ed. *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resuorces and Enviromental Sciences 20-23rd July 2010* University of Leicester, UK, 337– 340.

Ciepłuch, B., *et al.*, 2011. Assessing the quality of open spatial data for mobile location-based services research and applications. *Archives of photogrammetry, cartography and remote sensing, ISSN 2083-2214*, 22, 105–116.

Codescu, M., *et al.*, 2011. Osmonto-an ontology of openstreetmap tags. *State of the map Europe (SOTM-EU) 2011.*

Corcoran, P., Mooney, P., and Winstanley, A., 2010. Topological Consistent Generalization of OpenStreetMap. [online] [http://bit.ly/1U2OeyV].

Craglia, M., Ostermann, F., and Spinsanti, L., 2012. Digital Earth from vision to practice: making sense of citizen-generated content. *International Journal of Digital Earth*, 5 (5), 398–416.

Da Silva, A.C. and Wu, S.T., 2007. Consistent handling of linear features in polyline simplification. *In*: M.M. C.A. Davis Jr, ed. *Advances in Geoinformatics*. Springer, Switzerland, 1–17.

De Longueville, B., Ostländer, N., and Keskitalo, C., 2010. Addressing vagueness in Volunteered Geographic Information (VGI)–A case study. *International Journal of Spatial Data Infrastructures Research*, 5, 1725–0463.

De Tré, G., *et al.*, 2010. Consistently Handling Geographical User Data. *In*: F.H. E. Huellermeier R. Kruse, ed. *Information Processing and Management of Uncertainty in Knowledge-Based Systems, Applications, 28th June - 2nd July, 2010*. Springer, 85–94.

Fan, H., *et al.*, 2014. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28 (4), 700–719.

Flanagin, A.J. and Metzger, M.J., 2008. The credibility of Volunteered Geographic Information. *GeoJournal*, 72 (3-4), 137–148.

Foody, G., *et al.*, 2014. Accurate attribute mapping from volunteered geographic information: issues of volunteer quantity and quality. *The Cartographic Journal*, 52 (4).

Forghani, M. and Delavar, M.R., 2014. A quality study of the OpenStreetMap dataset for Tehran. *ISPRS International Journal of Geo-Information*, 3 (2), 750–763.

Frew, J., 2007. Provenance and volunteered geographic information. *Online available: http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Frew_paper.pdf*.

Fuchs, G., *et al.*, 2013. Tracing the German centennial flood in the stream of tweets: first lessons learned. *In*: A.V. D. Pfoser, ed. *Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, 5-8, November, 2013*, Orlando, FL, USA ACM, New York, NY, USA, 31–38.

Girres, J.F. and Touya, G., 2010. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14 (4), 435–459.

Golder, S.A. and Huberman, B.A., 2006. Usage patterns of collaborative tagging systems. *Journal of information science*, 32 (2), 198–208.

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

Goodchild, M.F. and Li, L., 2012. Assuring the quality of Volunteered Geographic Information. *Spatial statistics*, 1, 110–120.

Guinée, J.B., 2002. Handbook on life cycle assessment operational guide to the ISO standards. *The international journal of life cycle assessment*, 7 (5), 311–313.

Gupta, A. and Kumaraguru, P., 2012. Credibility ranking of tweets during high impact events. *In*: *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media, 17th April, 2012*, Lyon, France ACM, New York, NY, USA, p. 2.

Haklay, M., 2010. How good is Volunteered Geographic Information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning. B, Planning & design*, 37 (4), 682.

Haklay, M., *et al.*, 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.

Hartig, O., 2009. Provenance Information in the Web of Data.. *LDOW*, 538.

Hasan Dalip, D., *et al.*, 2009. Automatic quality assessment of content created collaboratively by web communities: a case study of wikipedia. *In*: *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries, 14-19 June, 2009*, Austin, TX, USA ACM, New York, NY, USA, 295–304.

Hashemi, P. and Abbaspour, R.A., 2015. Assessment of Logical Consistency in Open-

StreetMap Based on the Spatial Similarity Concept. *OpenStreetMap in GIScience*. Springer, 19–36.

Hecht, R., Kunze, C., and Hahmann, S., 2013. Measuring completeness of building footprints in OpenStreetMap over space and time. *ISPRS International Journal of Geo-Information*, 2 (4), 1066–1091.

Helbich, M., *et al.*, 2012. Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata. *Proceedings of GI_Forum, 3-6 July, 2012*, 24–33.

Hollenstein, L. and Purves, R., 2010. Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, (1), 21–48.

Hovland, C.I., Janis, I.L., and Kelley, H.H., 1953. *Communication and persuasion; psychological studies of opinion change.*. Yale University Press, New Haven, UK.

Hoyle, D., 2001. *ISO 9000: quality systems handbook*. Butterworth and Heinemann, Oxford, UK.

Huang, K.L., Kanhere, S.S., and Hu, W., 2010. Are you contributing trustworthy data?: the case for a reputation system in participatory sensing. *In*: *Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems, 17-21 October, 2010*, Bodrum, Turkey ACM, New York, NY, USA, 14–22.

Huberman, B.A., Romero, D.M., and Wu, F., 2008. Social networks that matter: Twitter under the microscope. *First Monday*, 14 (1).

Jackson, S.P., *et al.*, 2013. Assessing completeness and spatial error of features in volunteered geographic information. *ISPRS International Journal of Geo-Information*, 2 (2), 507–530.

Jacob, R., *et al.*, 2009. Campus guidance system for international conferences based on openstreetmap. *In*: G.T. J.M. Ware, ed. *Web and Wireless Geographical Information Systems*. Springer, Berlin Heidelberg, 187–198.

Jacobs, N., *et al.*, 2007. Geolocating static cameras. *In*: *IEEE 11th International Conference on Computer Vision (ICCV), 14-20 October, 2007*, Rio de Janeiro, Brazil IEEE, NJ, USA, 1–6.

Kang, B., O'Donovan, J., and Höllerer, T., 2012. Modeling topic specific credibility on twitter. *In*: *Proceedings of the ACM international conference on Intelligent User Interfaces, 14-17 February, 2012*, Lisbon, Portugal ACM, NJ, USA, 179–188.

Keßler, C. and de Groot, R.T.A., 2013. Trust as a proxy measure for the quality of Volunteered Geographic Information in the case of OpenStreetMap. *In*: J.C. D. Vandenbroucke B. Bucher, ed. *Geographic Information Science at the Heart of Europe*. Springer, Switzerland, 21–37.

Keßler, C., *et al.*, 2009. Bottom-up gazetteers: Learning from the implicit semantics of geotags. *In*: S.L. K. Janowicz M. Raubal, ed. *GeoSpatial semantics*. Springer, Berlin Heidelberg, 83–102.

Keßler, C., Trame, J., and Kauppinen, T., 2011. Tracking editing processes in volunteered geographic information: The case of OpenStreetMap. *In*: M.W. M. Duckham A. Galton, ed. *Identifying objects, processes and events in spatio-temporally distributed data (IOPE), workshop at conference on spatial information theory, 12-16 September, 2011*, Vol. 12, Belfast, Maine, USA.

Koukoletsos, T., Haklay, M., and Ellul, C., 2012. Assessing data completeness of VGI through an automated matching procedure for linear data. *Transactions in GIS*, 16 (4), 477–498.

Kounadi, O., 2009. Assessing the quality of OpenStreetMap data. *Thesis (MSc), Uni-*

versity College of London Department of Civil, Environmental And Geomatic Engineering.

Lindeberg, T., 2012. Scale invariant feature transform. *Scholarpedia*, 7 (5), 10491.

Liu, S.B., *et al.*, 2008. In search of the bigger picture: The emergent role of on-line photo sharing in times of disaster. *In*: B.V.d.W. F. Fiedrich, ed. *Proceedings of the Information Systems for Crisis Response and Management Conference (ISCRAM), 4-7 May, 2008*, Washington DC, USA.

MacEachren, A.M., *et al.*, 2011. Senseplace2: Geotwitter analytics support for situational awareness. *In*: *IEEE Conference on Visual Analytics Science and Technology (VAST), 23-28 October, 2011*, Providence, RI, USA, 181–190.

Maué, P., 2007. Reputation as tool to ensure validity of VGI. *In*: *Proceedings of the VGI specialist meeting, 13-14 December*, Santa Barbara, USA.

Chapter title. *The potential of citizen volunteered spatial information for building SDI* GSDI Association Press, 2009. .

Mendoza, M., Poblete, B., and Castillo, C., 2010. Twitter Under Crisis: Can we trust what we RT?. *In*: *Proceedings of the first workshop on social media analytics, 25-28 July*, Washington DC, USA ACM, New York, NY, USA, 71–79.

Milholland, N. and Pultar, E., 2013. The San Francisco public art map application: using VGI and social media to complement institutional data sources. *In*: *Proceedings of the 1st ACM SIGSPATIAL International Workshop on MapInteraction, 5-8 November*, Orlando, FL, USA, 48–53.

Montello, D.R., *et al.*, 2003. Where's downtown?: Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition & Computation*, 3 (2-3), 185–204.

Mooney, P. and Corcoran, P., 2012. The annotation process in OpenStreetMap. *Transactions in GIS*, 16 (4), 561–579.

Morris, M.R., *et al.*, 2012. Tweeting is believing?: understanding microblog credibility perceptions. *In*: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, 11-15 February*, Seattle, WA, USA, 441–450.

Mullen, W.F., *et al.*, 2015. Assessing the impact of demographic characteristics on spatial error in volunteered geographic information features. *GeoJournal*, 80 (4), 587–605.

Mummidi, L.N. and Krumm, J., 2008. Discovering points of interest from users? map annotations. *GeoJournal*, 72 (3-4), 215–227.

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

Neis, P., Zielstra, D., and Zipf, A., 2013. Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet*, 5 (2), 282–300.

O'Connor, R., 2009. GLOBAL: Facebook and Twitter 'reshaping journalism as we know it'. [online] Accessed: 2015-12-22 [http://bit.ly/1InIn6V].

O'Donovan, J., *et al.*, 2012. Credibility in context: An analysis of feature distributions in twitter. *In*: *Proceedings of the International Conference on Privacy, Security, Risk and Trust (PASSAT) and International Confernece on Social Computing (SocialCom), 3-5 September*, Amsterdam, Netherlands, 293–301.

Ostermann, F.O. and Spinsanti, L., 2011. A conceptual workflow for automatically assessing the quality of volunteered geographic information for crisis management. *In*: F.T. S. Geertman W. Reinhardt, ed. *Proceedings of the 14th AGILE Conference on Geographic Information Science, 18-21 April*, Utrecht, Netherlands.

Popescu, A., Grefenstette, G., and Moëllic, P.A., 2008. Gazetiki: automatic creation of a geographical gazetteer. *In*: *Proceedings of the 8th ACM/IEEE-CS joint conference* *on Digital libraries, 16-20 June*, Pittsburgh, PA, USA, 85–93.

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

Resnick, P., *et al.*, 2000. Reputation systems. *Communications of the ACM*, 43 (12), 45–48.

Robinson, S., *et al.*, 2012. Navigation your way: from spontaneous independent exploration to dynamic social journeys. *Personal and Ubiquitous Computing*, 16 (8), 973–985.

Sacha, D., *et al.*, 2016. The role of uncertainty, awareness, and trust in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 22 (1), 240–249.

Sakaki, T., Okazaki, M., and Matsuo, Y., 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. *In*: *Proceedings of the 19th international conference on World wide web, 26-30 April*, Raleigh, NC, USA, 851–860.

Schmitz, S., Neis, P., and Zipf, A., 2008. New applications based on collaborative geodatathe case of routing. *In*: *Proceedings of XXVIII INCA International Congress on Collaborative Mapping and Space Technology, 4-6 November*, Gandhinagar, Gujarat, India.

Senaratne, H., Bröring, A., and Schreck, T., 2013. Using Reverse Viewshed Analysis to Assess the Location Correctness of Visually Generated VGI. *Transactions in GIS*, 17 (3), 369–386.

Senaratne, H., *et al.*, 2014. Moving on Twitter: Using Episodic Hotspot and Drift Analysis to Detect and Characterise Spatial Trajectories. *In*: *7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN), 4-7 November*, Dallas, TX, USA ACM, New York, NY, USA.

Siebritz, L.A., Assessing the accuracy of openstreetmap data in south africa for the purpose of integrating it with authoritative data. Master's thesis, University of Cape Town, 2014. .

Tenney, M., 2014. Quality Evaluations on Canadian OpenStreetMap Data. *Spatial Knowledge and Information.*

Thomson, J., *et al.*, 2005. A typology for visualizing uncertainty. *In*: *Electronic Imaging 2005* International Society for Optics and Photonics, 146–157.

Valli, C. and Hannay, P., 2010. Geotagging Where Cyberspace Comes to Your Place.. *In*: *Proceedings of the International Conference on Security and Management (SAM '10), 12-15 July*, Las Vegas, NV, USA CSREA press, Athens, GA, USA, 627–632.

Van Exel, M., Dias, E., and Fruijtier, S., 2010. The impact of crowdsourcing on spatial data quality indicators. *Proceedings of GiScience 2011, 14-17 September.*

Van Oort, P. and Bregt, A., 2005. Do Users Ignore Spatial Data Quality? A Decision-Theoretic Perspective. *Risk analysis*, 25 (6), 1599–1610.

Vandecasteele, A. and Devillers, R., 2013. Improving Volunteered Geographic Data Quality Using Semantic Similarity Measurements. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1 (1), 143–148.

Vandecasteele, A. and Devillers, R., 2015. Improving Volunteered Geographic Information Quality Using a Tag Recommender System: The Case of OpenStreetMap. *OpenStreetMap in GIScience*. Springer, 59–80.

Wang, D., *et al.*, 2014. Using Semantic Techology for Consistency Checking of Road Signs. *In*: J.H.G.H. Z. Huang C. Liu, ed. *Web Information Systems Engineering–WISE 2013 Workshops, 13-15 October*, Nanjing, China Springer, Berlin Heidelberg, 11–22.

Will, J., 2014. Development of an automated matching algorithm to assess the quality of the OpenStreetMap road network: a case study in Göteborg, Sweden. *Student thesis series INES*.

Yanenko, O. and Schlieder, C., 2014. Game principles for enhancing the quality of user-generated data collections. *In*: *Proceeding of the 17th AGILE Workshop on Geogames Geoplay, 3-16 June*, Castellon, Spain, 1–5.

Zandbergen, P.A., Ignizio, D.A., and Lenzer, K.E., 2011. Positional accuracy of TIGER 2000 and 2009 road networks. *Transactions in GIS*, 15 (4), 495–519.

Zhang, W. and Kosecka, J., 2006. Image based localization in urban environments. *In*: *Third International Symposium on 3D Data Processing, Visualization, and Transmission, 14 -16 June*, Chapel Hill, North Carolina, USA.

Zielstra, D. and Hochmair, H.H., 2013. Positional accuracy analysis of Flickr and Panoramio images for selected world regions. *Journal of Spatial Science*, 58 (2), 251–273.