A conceptual model for quality assessment of VGI for the purpose of flood management

Lívia Castro Degrossi
University of São Paulo
São Carlos, Brazil
Heidelberg University
Heidelberg, Germany
degrossi@icmc.usp.br

João Porto de Albuquerque
University of Warwick
Warwick, England
GIScience Chair, Heidelberg University
Heidelberg, Germany
j.porto@warwick.ac.uk

Hongchao Fan, Alexander Zipf
GIScience Chair, Heidelberg University
Heidelberg, Germany
hongchao.fan@uni-heidelberg.de,
zipf@uni-heidelberg.de

Abstract

Volunteered Geographic Information (VGI) has emerged as a potential source of geographic information for different domains. Despite the many advantages associated with it, such information lacks of quality assurance, since it is provided by individuals with different motivations and backgrounds. In response to this, several methods have been proposed to assess the quality of volunteered geographic information of different platforms. However, there has been little investigation aimed at explaining how cross-platform data could be used for quality assessment. Moreover, it is not clear how the volunteer could be inserted in the quality assessment process in order to improve the overall quality. In this paper, we propose a conceptual model to assess the quality of volunteered geographic information for the purpose of flood management that combines cross-platform data, i.e. OpenStreetMap and social media data, and authoritative data. This model is part of an ongoing research for the development of an approach for quality assessment of VGI in a flood citizen observatory.

1 Introduction

Since last decade, Web 2.0 tools have changed the way information is created. Users now can play roles of both information consumer and provider. This change resulted in an unprecedented amount of information available online.

The term user-generated content refers to the information created by common individuals and distributed over the Internet [1]. A special case of user-generated content is an information with a geographic reference, also known as Volunteered Geographic Information (VGI) [2].

VGI is a collection of digital geographic information produced by ordinary individuals and informal institutions [2]. Like most Web 2.0 data, VGI is provided by individuals with little or no formal training in geo-spatial techniques [2, 3], different backgrounds and varying motivations to contribute data [4].

On the one hand, VGI has various advantages: (*i*) it is free; (*ii*) it is up-to-date; (*iii*) it provides different types of data (text, pictures, videos, etc.) [5]; (*iv*) it can be used to complement or substitute authoritative sources of information [6], etc. On the other hand, a core disadvantage of VGI is its lack of quality.

Concerns about its quality have emerged due to the vast amount of information provided by different individuals, creating great uncertainty regarding both who is responsible for the information and if source and information can be believed [7, 8]. These concerns are also related to the lack of quality control during the data creation process [7].

The lack of quality can affect the usability of VGI [9]. Hence, several methods have been proposed to evaluate VGI quality [10, 11, 12, 13, 14, 15], just to mention a few. However, no investigation has been carried out on how cross-linked VGI can be used to assess the quality of volunteered information in flood management domain. Moreover, in these methods, the volunteer is not part of the quality assessment process. Thus, it is necessary to investigate how the volunteer could be inserted into the process in order to improve the overall quality.

In this paper, we propose a conceptual model to assess the fitness of volunteered information for the purpose of flood management that combines cross-platform data (OpenStreetMap[1], Twitter[2], Instagram[3] and Flickr[4]) and authoritative data. This model is part of an ongoing research which aims to develop a quality assessment method for VGI in a flood citizen observatory.

The remainder of this paper is structured as follows: in Section 2, we discuss VGI quality and related works. In Section 3, there is a description of the conceptual model proposed in this work. Finally, Section 4 summarizes the conclusions of this work and future developments.

2 Quality of Volunteered Geographic Information

In the area of VGI, the issue of information quality is a core concern. The multiplicity of sources that ensure vast amount of information also leads to heterogeneous quality. Information quality comes with different definitions because it strongly depends on the application domain and the objectives of the application [16], i.e. it depends on the purpose for which the data will be used. This is referred to as “Fitness for

[1] http://openstreetmap.org/
[2] https://twitter.com/
[3] https://www.instagram.com/
[4] https://www.flickr.com/

Purpose". Hence, the assessment of VGI quality has to be adjusted to different use cases.

Ballatore and Zipf [17] proposed a framework to operationalize conceptual quality in VGI. This is an important concept to enable the semantic decoding of data, helping the consumer to decide if the information fits for her/his purpose. For OpenStreetMap data, Barron et al. [18] developed a framework for intrinsic analyses. The framework aims at facilitating the decision if the OSM dataset has sufficient quality for a use case or not.

Differently, a number of studies have been undertaken to evaluate quality in overall. A commonly used method is the comparison with authoritative data [11, 19, 20, 21]. However, this method can present some limitations associated with the reference dataset as financial costs and licensing restrictions [20], the dataset could have been obtained with old and less precise technologies [5] or the dataset could be updated in longer time periods [11]. To overcome these limitations, researchers have proposed new methods which do not require external data.

In different studies, data's historic are used as training set of automatic learning techniques. These techniques were applied to identify the assignment of wrong or implausible classes [10], to analyse semantic class information [14] and to estimate positional accuracy of VGI that have no corresponding reference data [15]. Another alternative is the VGI content itself. For visual VGI content, Senaratne et al. [22] developed a reverse viewshed method where VGI quality is determined by testing if the described object can be viewed from the position where the photo was geotagged. Another way is to measure the distance between the published image position and the estimated camera position based on image content [3].

Other methods take advantage of the crowd itself. Given a great number of participants, VGI quality can be determined based on a voting approach, where votes represent the popularity and quality of an information [23]. VGI quality can also be evaluated based on information from trusted sources: an example is the comparison with information provided by expert contributors [24, 25]. Finally, another way is based on the contributor's reputation. Bishr and Janowicz [9] proposed informational trust method as a social tie between a trustor and a trustee, i.e. it implies the transition of trust from the trustee to information entities conveyed by the trustee.

As can be seen, existing methods use authoritative data, metadata, or the crowd itself to assess VGI quality. However, there has been no investigation on how other sources of VGI – e.g. Twitter, Instagram, Flick, OpenStreetMap etc. – can be used in the quality assessment process. Moreover, these methods are not well-suited for the use in near-real time, which is, for example, an important requirement of VGI systems in the context of flood management. Finally, a common characteristic of these methods is the absence of a feedback, about the information quality, to the contributor. Thus, he/she does not know if the provided information can be used or not.

3 Conceptual Model

The quality of volunteered geographic information has gained special attention due to the growing use of VGI in different application domains. Hence, how to assess its quality has become the subject of several studies as presented in the previous section. However, previous studies still do not explore the potential of cross-linked VGI during the quality assessment process.

To overcome this gap, we propose a conceptual model to assess VGI quality. As illustrated in Figure 1, the model consists of four main steps. We use a flood scenario to demonstrate and discuss the potentials of the proposed model.

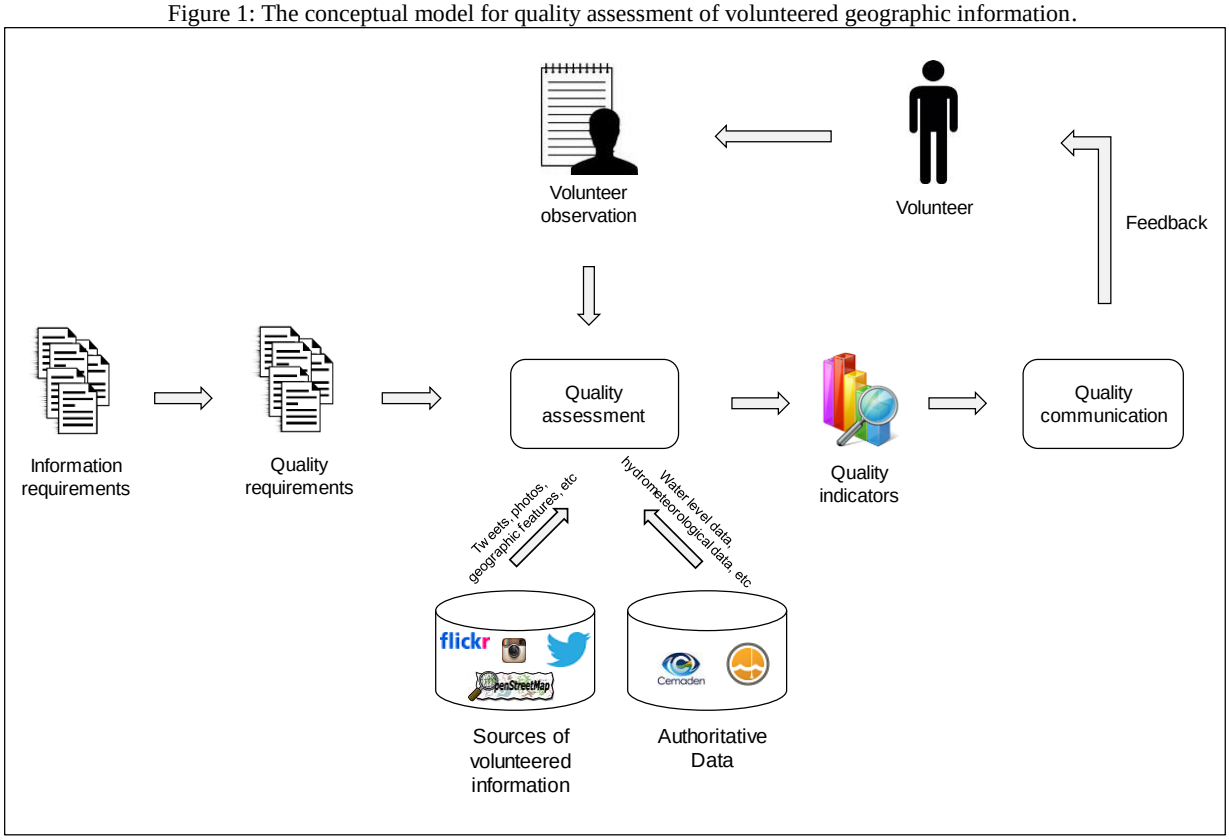
The first step is the definition of information requirements. Depending on the application domain, different information requirements can be identified. In this work, a set of information requirements was derived based on a literature review. The requirements were validated through a survey with specialists (hydrologists, decision makers etc.). The survey aimed at answering if the set was in accordance with the information used in flood management domain.

The second step corresponds to the definition of quality elements for each information requirement. A quality element is "a component describing a certain aspect of the quality of geographic data" [26]. The International Organization for Standardization (ISO) defines five quality elements: completeness, thematic accuracy, logical consistency, temporal quality and positional accuracy [26]. We used these elements to describe the quality of a volunteered observation. To the best of the authors' knowledge, there is no systematic method to derive quality elements for geographic information. Hence, we derived quality element(s) for each information based on information type and attributes, and the objective of the application domain.

Once the quality elements have been defined, we used them as a basis to derive quality indicators (parameters) from social media and OpenStreetMap data, in order to measure the quality of volunteered information in citizen observatories. We use social media data because "social media message can provide information about what is happening when and where" [27]. It is important to highlight that the resulting quality indicators will be validated against authoritative data.

The third step comprises the quality assessment itself. For this, the quality of each volunteer observation is measure based on the quality indicators. An example is given to a better understanding of this step. In a flood scenario, contributors usually provide observations about the imminence or occurrence of a flood. Whether the event can be identified simultaneously in different VGI sources, it indicates that the event is probably happening and the observation is possibly true. It has been demonstrated, for example, that the periods when the traditional media talks about floods in the United Kingdom correspond to the periods when contributors most intensively upload flood-related photos in Flickr [28]. Moreover, disaster-related messages containing images, e.g. messages from Instagram and Flickr, tend to be closer to the event [27]. Finally, a spatial analysis of tweets after the River Elbe Flood of June 2013 in Germany showed that compared with the overall flood-related tweets, "there is perhaps a tendency for 'relevant' on-topic tweets to be closer to flood-affected catchments" [29].

Figure 1: The conceptual model for quality assessment of volunteered geographic information.



An important aspect of a flood is its location because floods are local events, i.e. they occur in a specific area. If a contributor provides an information about a flood, then it is expected that the location of the information is close to a water resource. Hung et al. [13], for example, demonstrated that reports with high credibility are mostly located in flood-risk areas, indicating that these areas would be an important factor in determining credibility. However, if flood-risk data is not available, they suggest using distance to water resources areas as a proxy measure. In order to evaluate VGI location, we propose to use OpenStreetMap (OSM) data as a source of geographic features.

Finally, the fourth step is the communication of the quality indicator(s) to the volunteer. This step aims to provide a feedback about the information quality to the volunteer and how to improve it. It is based on the assumption that when the volunteer knows the information quality and learns how to improve it, he/she can provide information with higher quality.

4 Summary and future developments

Based on the nature of VGI, the information is provided by ordinary citizens which would lead to varying quality and possible data inconsistencies. Because of this, it is important to evaluate the quality of such information before it is used in different application domains. Here, we proposed a conceptual model for the quality assessment of volunteered geographic information that combines cross-linked VGI and authoritative data.

The model has been applied to assess VGI quality in a disaster context, more specifically floods. Until now, we have identified what the information requirements are and the quality elements corresponding to each information. The next stage in this research is to define quality indicators based on social media and OpenStreetMap data. These indicators will be used to measure VGI quality. Finally, an evaluation will be carried out in order to verify the consistency of such indicators, i.e. for similar information, it is expected similar quality indicators, and possible limitations.

References

[1] T. Daugherty, M. S. Eastin, and L. Bright, “Exploring consumer motivations for creating user-generated content,” *J. Interact. Advert.*, vol. 8, no. 2, pp. 16–25, 2008.

[2] M. F. Goodchild, “Citizens as sensors: the world of volunteered geography,” *GeoJournal*, vol. 69, no. 4, pp. 211–221, Aug. 2007.

[3] D. Zielstra and H. H. Hochmair, “Positional accuracy analysis of Flickr and Panoramio images for selected world regions,” *J. Spat. Sci.*, vol. 58, no. 2, pp. 251–273, 2013.

[4] M. Bishr and K. Janowicz, “Can we Trust Information ? - The Case of Volunteered Geographic Information,” in *Proceedings of the Workshop Towards Digital Earth: Search, Discover and Share Geospatial Data*, 2010.

[5] M. F. Goodchild and L. Li, “Assuring the quality of volunteered geographic information,” *Spat. Stat.*, vol. 1, pp. 110–120, 2012.

[6] R. Karam and M. Melchiori, “A Crowdsourcing-Based Framework for Improving Geo-spatial Open Data,” in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2013, pp. 468–473.

[7] A. J. Flanagin and M. J. Metzger, “The credibility of volunteered geographic information,” *GeoJournal*, vol. 72, no. 3–4, pp. 137–148, Jul. 2008.

[8] G. M. Foody, L. See, S. Fritz, M. Van der Velde, C. Perger, C. Schill, and D. S. Boyd, “Assessing the Accuracy of Volunteered Geographic Information arising from Multiple Contributors to an Internet Based Collaborative Project,” *Trans. GIS*, vol. 17, no. 6, pp. 847–860, 2013.

[9] M. Bishr and W. Kuhn, “Geospatial Information Bottom-Up: A Matter of Trust and Semantics,” in *The European Information Society*, 2007, pp. 365–387.

[10] A. L. Ali and F. Schmid, “Data Quality Assurance for Volunteered Geographic Information,” in *Geographic Information Science*, vol. 8728, Springer International Publishing, 2014, pp. 126–141.

[11] H. Fan, A. Zipf, Q. Fu, and P. Neis, “Quality assessment for building footprints data on OpenStreetMap,” *Int. J. Geogr. Inf. Sci.*, vol. 28, no. 4, pp. 700–719, 2014.

[12] G. M. Foody, L. See, S. Fritz, M. Van der Velde, C. Perger, C. Schill, and D. S. Boyd, “Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project,” *Trans. GIS*, vol. 17, no. 6, pp. 847–860, 2013.

[13] K.-C. Hung, M. Kalantari, and A. Rajabifard, “Methods for assessing the credibility of volunteered geographic information in flood response: A case study in Brisbane, Australia,” *Appl. Geogr.*, vol. 68, pp. 37–47, Mar. 2016.

[14] M. Jilani, P. Corcoran, and M. Bertolotto, “Automated Highway Tag Assessment of OpenStreetMap Road Networks,” in *22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2014.

[15] N. Mohammadi and M. Malek, “Artificial intelligence-based solution to estimate the spatial accuracy of volunteered geographic data,” *J. Spat. Sci.*, vol. 60, no. 1, pp. 119–135, 2015.

[16] G. Bordogna, P. Carrara, L. Criscuolo, M. Pepe, and A. Rampini, “On predicting and improving the quality of Volunteer Geographic Information projects,” *Int. J. Digit. Earth*, vol. 0, no. 0, pp. 1–22, 2014.

[17] A. Ballatore and A. Zipf, “12th International Conference Spatial Information Theory (COSIT 2015),” Cham: Springer International Publishing, 2015, pp. 89–107.

[18] C. Barron, P. Neis, and A. Zipf, “A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis,” *Trans. GIS*, vol. 18, no. 6, pp. 877–895, 2014.

[19] J. F. Girres and G. Touya, “Quality Assessment of the French OpenStreetMap Dataset,” *Trans. GIS*, vol. 14, no. 4, pp. 435–459, 2010.

[20] P. Mooney, P. Corcoran, and A. C. Winstanley, “Towards quality metrics for OpenStreetMap,” in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, pp. 514–517.

[21] K. Poser and D. Dransch, “Volunteered Geographic Information for Disaster Management with Application to Rapid Flood Damage Estimation,” *Geomatica*, vol. 64, no. 1, pp. 89–98, 2010.

[22] H. Senaratne, A. Bröring, and T. Schreck, “Using Reverse Viewshed Analysis to Assess the Location Correctness of Visually Generated VGI,” *Trans. GIS*, vol. 17, no. 3, pp. 369–386, 2013.

[23] V. Lertnattee, S. Chomya, and V. Sornlertlamvanich, “Reliable Improvement for Collective Intelligence on Thai Herbal Information,” *Stud. Comput. Intell.*, vol. 283, pp. 99–110, 2010.

[24] L. See, A. Comber, C. Salk, S. Fritz, M. van der Velde, C. Perger, C. Schill, I. McCallum, F. Kraxner, and M. Obersteiner, “Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts,” *PLoS One*, vol. 8, no. 7, pp. 1–11, 2013.

[25] A. Comber, L. See, S. Fritz, M. Van der Velde, C. Perger, and G. Foody, “Using control data to determine the reliability of volunteered geographic information about land cover,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 23, p. 37-48, 2013.

[26] ISO 19157:2013, “Geographic information - Data quality (ISO 19157:2013).” Brussels, Belgium, 2013.

[27] R. Peters, “Exploring the role of image-based social media messages for disaster management: Comparing various social media platforms in the case of the Elbe flood 2013 in Germany.” 2015.

[28] B. De Longueville, G. Luraschi, P. Smits, S. Peedell, and T. De Groeve, “Citizens as Sensors for Natural Hazards: A VGI integration Workflow,” *Geomatica*, vol. 64, no. 1,

pp. 41–59, 2010.

[29] J. P. Albuquerque, B. Herfort, A. Brenning, and A. Zipf, "A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management," *Int. J. Geogr. Inf. Sci.*, vol. 29, no. 4, pp. 667–689, 2015.