

ASSESSING VOLUNTEERED GEOGRAPHIC INFORMATION (VGI) QUALITY BASED ON CONTRIBUTORS' MAPPING BEHAVIOURS

D. Bégin ^{a,b,*}, R. Devillers ^{a,b}, S. Roche ^b

^a Department of Geography, Memorial University, St. John's (NL), Canada - (d.begin, rdeville)@mun.ca

^b Centre de recherche en géomatique, Université Laval, Québec (QC), Canada - stephane.roche@scg.ulaval.ca

KEY WORDS: Volunteered Geographic Information (VGI), OpenStreetMap (OSM), Contributors Behaviour, Spatial Data Quality Assessment, Data Completeness

ABSTRACT:

VGI changed the mapping landscape by allowing people that are not professional cartographers to contribute to large mapping projects, resulting at the same time in concerns about the quality of the data produced. While a number of early VGI studies used conventional methods to assess data quality, such approaches are not always well adapted to VGI. Since VGI is a user-generated content, we posit that features and places mapped by contributors largely reflect contributors' personal interests. This paper proposes studying contributors' mapping processes to understand the characteristics and quality of the data produced. We argue that contributors' behaviour when mapping reflects contributors' motivation and individual preferences in selecting mapped features and delineating mapped areas. Such knowledge of contributors' behaviour could allow for the derivation of information about the quality of VGI datasets. This approach was tested using a sample area from OpenStreetMap, leading to a better understanding of data completeness for contributor's preferred features.

1. INTRODUCTION

Developments in mobile location technologies and Web applications have increased the number of people creating and sharing geographic information over the Web (van Exel and Dias, 2011). This has allowed communities of users to develop around collaborative online mapping projects such as OpenStreetMap (OSM) (Haklay and Weber, 2008). In a number of contexts, this so called 'volunteered geographic information' (VGI) has provided richer and more up-to-date geographic data than national mapping agencies (NMAs) or other authoritative data sources (Mashhadi et al., 2012; Mooney et al., 2011).

While many organizations could benefit from VGI, only a few are using such data due to, amongst other reasons, a lack of reliable methods for assessing the quality of VGI data for an area of interest (van Exel et al., 2010). Attempts to assess VGI quality have highlighted data heterogeneity as an intrinsic characteristic of these data (Girres and Touya, 2010; Haklay et al., 2010; van Exel et al., 2010).

VGI data heterogeneity is a direct consequence of the collaborative processes used to produce these maps (Bruns, 2006; Goodchild, 2007). Studies have proposed describing VGI users' contributions using three components: the motivation, the action and the outcome (Budhathoki et al., 2010; Rehrl et al., 2013). Using this framework, features and places mapped by contributors (action) are believed to largely reflect contributors' personal interests (motivation) that result in heterogeneous datasets (outcome).

This paper proposes deriving the quality of VGI datasets from an understanding of contributors' behaviour when mapping. Section 2 will present an overview of current approaches in assessing VGI data quality. Section 3 will propose a new approach for assessing VGI data based on contributors' behaviour. Section 4 will highlight the potential impacts of the

method on evaluating VGI data completeness. Finally, Section 5 will illustrate the approach using a small test area with OSM data.

2. QUALITY ASSESSMENT OF VGI DATA

Chrisman (2006) traces the roots of map accuracy standards and spatial data quality assessment methods to early NMAs work. These methods were developed to ensure that mapping operations provided maps of consistent quality over large territories (Harding, 2006).

Spatial data quality can be described using different quality elements, including positional, thematic and temporal accuracies, logical consistency, and completeness (ISO/TC 211, 2002). While studies have used those elements for characterizing the quality of VGI data (Girres and Touya, 2010; Haklay, 2010; Neis et al., 2011), other quality elements specific to VGI are likely to be required. New VGI assessment methods will have to overcome some of the limitations of traditional approaches (van Exel et al., 2010).

2.1 Conventional approaches, a first look at VGI

VGI changed the mapping landscape by allowing people that are not professional cartographers to contribute to large and complex mapping projects. As a consequence, an initial concern with VGI was assessing users' credibility and understanding users' motivation (Coleman et al., 2009; Flanagan and Metzger, 2008). Some of these concerns became secondary with the publication of early data quality assessment studies comparing OpenStreetMap data, one of most successful VGI projects (Haklay and Weber, 2008), with authoritative datasets from different countries. Studies assessed the positional accuracy and completeness of OSM road network (Haklay, 2010; Zielstra and Zipf, 2010) and natural features (Mooney et al., 2010). Girres

* Corresponding author.

and Touya (2010) provided a comprehensive study of OSM quality elements for both road network and natural features.

These early studies found three characteristics of VGI data. First, the positional accuracy of VGI can be very high for man-made features, while accuracy proved to be lower for natural features. Second, VGI data usually proved to be accurate (Haklay, 2010; Mashhadi et al., 2012), even considering semantic differences between OSM and authoritative datasets used to assess VGI data (Al-Bakri and Fairbairn, 2012). Third, while VGI data quality is high in populated places, the quality of the data varies spatially and suffers from completeness problems in less populated places.

Recognizing the difficulty in assessing VGI data quality using conventional assessment methods, studies proposed new approaches to assess the quality of VGI data and mitigate the effects of data heterogeneity. These approaches are classified here using two categories proposed by Goodchild and Li (2012).

2.2 Crowdsourcing approaches: the effect of the number

Crowdsourcing approaches are data-centric methods that aim at using metadata as proxy measures for data quality. Several studies using such an approach explored the relationship between contributors' density and data quality. Haklay et al. (2010) proposed using the Linus's law (Raymond, 1999) as a framework to study VGI data quality. Studies have confirmed a relationship between the density of contributors and the quality of the data (Haklay et al., 2010; Napolitano and Mooney, 2012; Neis et al., 2011; Neis and Zipf, 2012), although the nature of this relationship is not yet clearly understood.

Other authors have looked at the relationship between the number of edits and the quality of the features (Keßler et al., 2011; Mashhadi et al., 2012; Mooney and Corcoran, 2012a). In these cases, data analyses proved that "the quality of contributions in OSM is independent of the number of edits/revisions they have undergone" (Mashhadi et al., 2012).

2.3 Social behaviour approaches: who should be trusted?

A last type of approach uses the level of trust one has in individual contributors to assess the quality of the data they produced. Trust levels are widely used and studied as a quality metric for Web sites (Adler and De Alfaro, 2006; Bishr and Kuhn, 2007; Javanmardi et al., 2010) and were proposed as a "people-object transitivity of trust" for VGI (Bishr and Janowicz, 2010).

Two approaches were proposed to assess VGI contributors' trust level. The first one is data-centric and relies on features' editing history (Keßler et al., 2011; Mashhadi et al., 2012; Mooney and Corcoran, 2012b). The second one proposes combining data-centric and user-centric elements to better quantify "the collective intelligence of the crowd generating data" (van Exel et al., 2010; van Exel and Dias, 2011). Both approaches seem promising but results did not provide a clear measure that could be used as a proxy for data quality.

3. VGI CONTRIBUTORS' MAPPING BEHAVIOUR

VGI data quality assessment studies have highlighted differences between two worlds. The first one is the world of

industrial mapping, providing consistent and uniform contents. The second one is a world of individuals, providing highly heterogeneous contents in a perpetual 'work in progress' map.

Unlike conventional cartographic data, crowdsourcing geographic knowledge leads to highly heterogeneous data (Feick and Roche, 2013) described as a "patchwork of geographic information" (Goodchild, 2007), and seen as a paradigm shift from a "layer cakes" (data layers of full coverage) to a "cupcakes" (local detailed contributions) view of the world (Roche, 2012).

Since VGI is user-generated content (Bruns, 2006), the features and places mapped by contributors and the sequence in which they were created are believed to largely reflect contributors' personal interests and behaviour. A VGI dataset is then a collection of contributions from multiple individuals that could be influenced by individuals' spatial preferences, feature type preferences, and mapping behaviours. This paper suggests studying contributors' preferences and mapping behaviour to understand the characteristics of the data produced and derive information about data quality.

3.1 Contributors' participation as mapping processes

VGI relies on contributors' participation. A characteristic of participation in online communities is that the frequency of contributions can vary by orders of magnitude between users. This has been described as the '90-9-1' rule by Nielsen (2006) and relates to the Zipf's law distribution (Li, 2002; Wyllys, 1981).

In the case of OpenStreetMap, Neis and Zipf (2012) have shown that about 5% of all OSM registered users have produced almost 90% of the transactions (changesets). Similarly, Mooney and Corcoran (2012b) found that only 20 contributors created about 61% of all 'ways' in London. As a consequence, studying the contribution of few major contributors of a dataset should provide information on most of its content.

Deriving information on data quality from the study of individual contributors' mapping behaviour also has the advantage of being broadly usable. The knowledge of a contributors behaviour gained in one region could be used to assess data created by this contributor in another region.

3.2 Feature type preferences

We posit that VGI contributions are largely influenced by individual contributors' interests. For instance, while some contributors may mostly map roads, others prefer to focus on hiking trails. Personal preferences and mapping behaviours can also influence the sequence in which contributors will map features (e.g., start mapping roads and then map main buildings). We define 'feature type preferences' as the inclination of a contributor to capture most instances of a specific feature type (e.g., main roads, buildings) before capturing features of lower priority (e.g., secondary roads, parks) within the same mapping area. Some feature types can even be systematically ignored by some contributors. Feature type preferences can be identified as 'pet features', in an analogy to the concept of 'pet location' proposed by Napolitano and Mooney (2012) that refers to locations of particular interest for contributors.

3.3 Mapping area selection and delineation

Napolitano and Mooney (2012) have shown that ‘dedicated’ OSM contributors have ‘pet locations’, usually areas contributors know well (e.g., close to home or work). Such local knowledge was proposed as a proxy measurement for data quality as the quality of the data is expected to be higher for areas contributors know best (van Exel et al., 2010).

Our approach assumes that a contributor will not randomly select objects and location to map but will rather tend to orderly map features based on personal preferences and on spatial adjacency. Figure 1 provides an example of a VGI contributor that maps a new area in successive editing sessions, completing a different changeset each time.

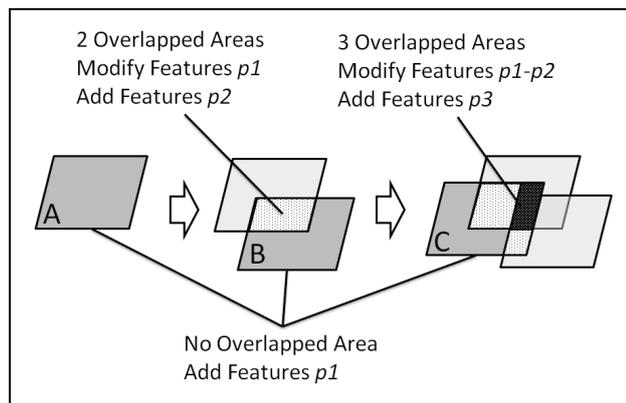


Figure 1: Simplified mapping process history and corresponding feature edition. Areas A, B and C represent three successive editing sessions (i.e., changesets). Feature types $p1$, $p2$ and $p3$ are higher to lower priority features types.

In a first session (*Changeset A*), the contributor will create features of higher preference ($p1$). In a second session (*Changeset B*), the contributor can: (1) extend spatially the mapping of higher preference features ($p1$), (2) modify $p1$ features in previously mapped area or (3) map lower preference features ($p2$) in area where $p1$ features were mapped. The last two operations occur where the current changeset overlaps previous ones. In the following sessions (*changeset C*), the contributor’s editing possibilities depend on the number of overlapping changesets (N): the contributor can extend the mapping of *higher preference* features ($p \leq N$), modify *higher preference* features ($p < N$) in areas mapped previously, or map lower preference features ($p \leq 2-N$) in areas where *higher preference* features were mapped.

A contributor is expected to extend the mapping area within/between editing sessions (*changesets*), until the contributor has no more interest in expanding the area for a specific feature type. In practice, multiple contributors will often be involved in mapping an area. Each contributor will take into account the features provided by the others, considering his own preferences. Understanding contributors’ implicit mapping processes is expected to help defining the extension of mapped areas and thereby data completeness, a key data quality element for VGI data.

4. ASSESSING VGI DATA COMPLETENESS

In conventional data quality assessments, ‘data completeness’ is a data quality element describing the “presence and absence of features, their attributes and relationships” (ISO/TC 211, 2002). In practice, VGI users are likely more concerned by missing data (omission) than data that are in excess (commission). VGI being a map in constant progress, no metadata describe an area as being completely mapped for a specific feature. Hence, VGI data completeness assessments should describe both the proportion of features omission in mapped area and the proportion of unmapped features in unmapped areas.

In this context, defining the extent of area mapped by contributors is key in this approach. The OSM community uses minimum bounding rectangles (MBR), also called bounding boxes, to define the area covered by individual mapping sessions. Since MBRs do not take into account the distribution of edits within their boundaries, we propose using concave hulls to define mapped areas for each editing session. Concave hulls have the advantage to minimize inclusion of empty areas compared to convex hulls or MBRs. Furthermore, we posit that contributors’ feature type preferences allow discriminating between area where higher priority features are expected to be mapped, and area where they are not expected to be mapped. Assuming higher priority features are mapped first ($p1$), the mapping of lower priority features ($p2$, $p3$) by a contributor could indicate that higher priority features were completed in the given area, resulting in knowledge of features completeness for $p1$ in this area.

Contributors’ mapped areas defined using concave hulls could allow the discrimination between mapped and unmapped areas within a same dataset. Characterizing the preferences in feature types and the area mapped by the main contributors could help define where VGI data quality assessment can be done, and on which features it can be done.

5. ILLUSTRATION OF THE APPROACH

A preliminary analysis was conducted over a sample area as a proof of concept of the proposed approach. The analysis had for objective to confirm if contributors do display feature type preferences and to explore how knowing those preferences can help assessing data completeness for given areas. A 4 km² area covering both urban and rural environments was selected in the vicinity of St. John’s, Newfoundland and Labrador, Canada. Data and metadata were obtained using the standard OSM web application programming interface (OpenStreetMap Wiki, 2013). Data manipulation and analysis were performed using FME software*.

5.1 Contributors’ edits in test area

The main contributors for this area were identified using a frequency count of the ‘user’ key values over all components of the dataset (i.e., nodes, ways and relations).

As predicted by Zipf’s law, it was found that three of the 17 current contributors have mapped or updated almost 95% of the area (Figure 2). The three main OSM contributors are the users *bgamberg*, *jfd553* and *cicerone*.

* <http://www.safe.com/>

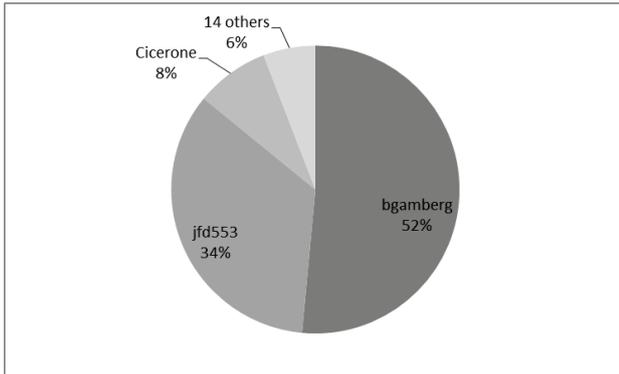


Figure 2: Percentage of contributions of individual OSM users in the test area.

5.2 Feature type preferences

The OSM schema does not differentiate between features and attributes but rather associates ‘key = value’ tuples named ‘tags’ to geometric objects (i.e., a residential street can be described using tags: highway = trunk and name = Graham. We used the ‘key’ component to identify the ‘map features’ prescribed by the community (OpenStreetMap Wiki, 2012).

Since 2009, OSM stores all edits made to the map using changesets (OpenStreetMap Wiki, 2013). Each changeset is the result from an editing session. Features mapped by the main three contributors were identified from their editing history. All changesets of the three main contributors were downloaded for analysis. Results indicate that contributors have a combined contribution of almost 20 million edits. A more detailed analysis has shown that two of them have made major data imports from Canadian NMA datasets. For the purpose of the study, these imports were removed from our analysis as they are not indicative of the contributors’ behaviour discussed earlier. About 281,000 edits were kept for analysis.

We used the frequency count of each feature type mapped by main contributors as a proxy measure of contributor’s feature type preferences. Results of the analysis show large differences in contributors’ interests (Figure 3). If the most popular feature is the road network (i.e., ‘highway’ key), it represents 95% of all *cicerone*’s contributions but less than 10% for *bgamberg*. Natural features represent about 85% of *bgamberg*’s contributions but less than 20% for *jfd553* and almost nothing for *cicerone*.

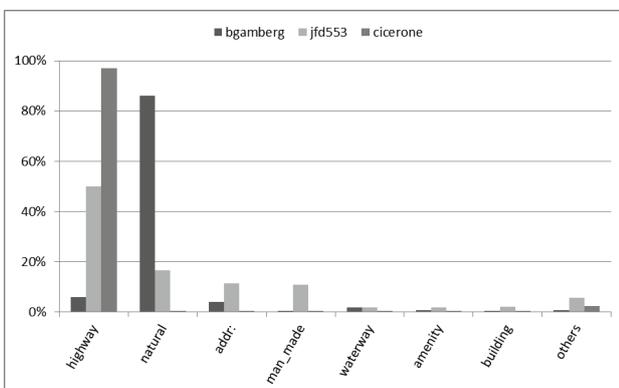


Figure 3 Percentages of features edited by each of the three main contributors.

The selection of other features to map varies amongst contributors, showing their personal preferences. These more specific features are often related to contributors’ local knowledge, as described by Napolitano and Mooney (2012).

5.3 Mapping area selection and delineation

The distribution of mapping areas of the three main contributors was also analysed using editing histories. The analysis of contributions revealed that one of the main contributors has mapped in over 30 countries, while the other two have only mapped in four neighbouring countries. To understand contributors’ mapping behaviour, we selected one location per contributor outside of the test area. We analysed the entire editing history of these areas in order to understand how edits made by other contributors influenced main contributors’ behaviour. We found that edits made by our three main contributors usually adhered to the previously identified individual feature type preferences in our test area (Figure 3).

For each area, editing sessions (changesets) of main contributors were delimited using concave hulls. As expected, new changesets usually extend or overlap spatially earlier changesets, creating a growing patchwork of edits. Furthermore, we also confirmed, at least for our main contributors, that overlapped edits usually add lower priority features or new attributes to existing higher priority features (Figure 4). In this example, a contributor has mainly created ‘highway’ features in changeset A, ‘amenity’ and ‘addresses’ in changeset B, and ‘footways’ in changeset C.

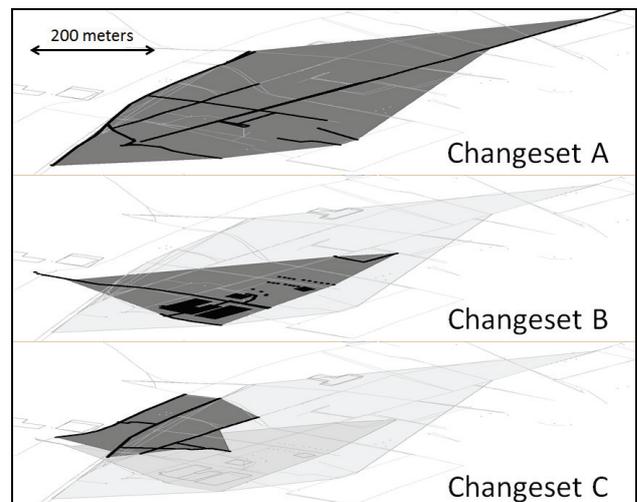


Figure 4: Overlapping concave hulls showing the spatial extent of three sequential editing sessions (changesets A, B and C). Black lines and polygons are new features edited by a contributor. Dark areas represent current changeset and lighter grey areas are previous changesets.

5.4 Proof of concept using a test area

As expected, contributors’ behaviour analysed in other regions proved to be similar to behaviours observed in the test area.

The overlap of the main contributors’ concave hulls allowed identifying unmapped areas. These unmapped areas were rural areas where there were little or no man-made features to map. Some areas were covered by more than 15 hulls, indicating areas of high mapping activity by the contributors.

All the road network was covered by at least four hulls, indicating that contributors have mapped at least four times in the vicinity, thus increasing the likelihood that all roads were mapped. A data completeness validation for contributor's feature type preferences (highway and natural features) was achieved using photo interpretation and field completion in the test area.

Results show that most of the road network was mapped properly with the exception of service roads. In this case, the analysis shows that contributors do not only have priorities in terms of features, but also within individual feature types. For instance, contributors usually mapped roads of higher importance before mapping residential roads.

Similar observations were made with natural features. While the entire hydrological network was mapped, none of the wooded areas were. Looking at specific preferences of main contributors for natural features, we found that two contributors (*jfd553* and *bgamberg*) prioritized water bodies over wooded areas. The last contributor (*cicerone*) has even mapped one of the few water bodies ever contributed.

Some contributors' lower priority features were mapped but their distribution was uneven, creating small clusters within the test area. The configuration and the content of these clusters suggest that contributors had specific local knowledge of and interest in these areas. The mapping of features that need local knowledge proves to be more difficult to predict based on contributor's preferences.

6. CONCLUSION

Assessing the quality of VGI data using conventional assessment methods proves to be difficult. Alternative approaches have been proposed but current results do not suggest any clear measurement that could be used as a proxy for data quality. This paper suggested studying contributors' mapping behaviour, such as preferences for objects to map, to understand the characteristics and quality of the data produced.

The advantage of the proposed approach is threefold. First, based on the Zipf's law, analysing the behaviour of only few contributors could allow the characterization of most datasets. Second, since contributors seem to have feature types preferences ('pet features'), it is possible to predict the features that are likely to be created by a contributor in a dataset, and the features that may not be created. Third, based on feature types preferences, a contributor will tend to map adjacent priority features in order to complete the mapping of a location of interest. Using concave hulls in defining each editing sessions, it was possible to produce a detailed image of the main contributors' mapped area.

The approach was tested using a sample dataset. The analysis of contributors' edits provided insights on their feature type preferences and mapping behaviour, allowing in turn to get an indication of data completeness for highest priority features in a VGI dataset. Missing data were related to the existence of lower priority levels within a feature type (e.g., primary vs. service roads).

However, analyses for larger and more diverse areas will have to be conducted in order to better assess the potential of the method for deriving information about data completeness.

Future works will also be looking at the impact of priorities, within feature types, on data completeness predictions and will examine how a better definition of mapped area using concave hulls could help refining data quality assessment in VGI.

ACKNOWLEDGEMENTS

The authors would like to thankfully acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support.

REFERENCES

- Adler, B.T., De Alfaro, L., 2006. *A content-driven reputation system for the Wikipedia*, Technical Report ucsc-crl-06-18, Santa Cruz (USA).
- Al-Bakri, M., Fairbairn, D., 2012. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. *International Journal of Geographical Information Science*, 26(8), pp. 1437-1456.
- Bishr, M., Janowicz, K., 2010. Can we trust information?-the case of volunteered geographic information, *In: Towards Digital Earth Search Discover and Share Geospatial Data Workshop at Future Internet Symposium*, Berlin (DEU), 20 September 2010, 6 p.
- Bishr, M., Kuhn, W., 2007. Geospatial information bottom-up: A matter of trust and semantics. *The European information society*, pp. 365-387.
- Bruns, A., 2006. Towards Producers: Futures for user-led content production, *In: Proceedings of Cultural Attitudes towards Communication and Technology*, Tartu (EST), 28-01 July 2006, pp. 275-284.
- Budhathoki, N.R., Nedovic-Budic, Z., Bruce, B., 2010. An interdisciplinary frame for understanding volunteered geographic information. *Geomatica*, 64(1), pp. 11-26.
- Chrisman, N., 2006. Development in the treatment of spatial data quality, *Fundamentals of spatial data quality*, Devillers, R. and Jeansoulin, R. (Eds.), ISTE, London (GBR), pp. 21-30.
- Coleman, D.J., Georgiadou, Y., Labonté, J., 2009. Volunteered geographic information: The nature and motivation of producers. *International Journal of Spatial Data Infrastructures Research*, 4, pp. 332-358.
- Feick, R., Roche, S., 2013. Understanding the Value of VGI, *Crowdsourcing Geographic Knowledge: VGI in Theory and Practice*, Sui, D.Z., Elwood, S. and Goodchild, M.F. (Eds.), Springer, Netherlands, pp. 15-29.
- Flanagin, A.J., Metzger, M.J., 2008. The credibility of volunteered geographic information. *GeoJournal*, 72(3-4), pp. 137-148.
- Girres, J., Touya, G., 2010. Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, 14(4), pp. 435-459.

- Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), pp. 211-221.
- Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning & design*, 37(4), pp. 682-703.
- Haklay, M., Basiouka, S., Antoniou, V., Ather, A., 2010. How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *The Cartographic Journal*, 47(4), pp. 315-322.
- Haklay, M., Weber, P., 2008. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing*, 7(4), pp. 12-18.
- Harding, J., 2006. Vector Data Quality: A Data Provider's Perspective, *Fundamentals of spatial data quality*, Devillers, R. and Jeansoulin, R. (Eds.), ISTE, London (GBR), pp. 141-159.
- ISO/TC 211, 2002. *19113 Geographic Information - Quality Principles*, 1st ed. International Organization for Standardization (ISO), Geneva (CHE).
- Javanmardi, S., Lopes, C., Baldi, P., 2010. Modeling user reputation in wikis. *Statistical Analysis and Data Mining*, 3(2), pp. 126-139.
- Kebler, C., Trame, J., Kauppinen, T., 2011. Tracking editing processes in volunteered geographic information: The case of OpenStreetMap, *In: Proceedings of Workshop on Identifying Objects, Processes and Events in Spatio-Temporally Distributed Data (IOPE)*, Belfast (USA), 12 September 2011, 7 p.
- Li, W., 2002. Zipf's law everywhere. *Glottometrics*, 5, pp. 14-21.
- Mashhadi, A., Quattrone, G., Capra, L., Mooney, P., 2012. On the Accuracy of Urban Crowd-Sourcing for Maintaining Large-Scale Geospatial Databases, *In: 8th International Symposium on Wikis and Open Collaboration - WikiSym '12*, Linz (AUT), August 27-29 2012, 10 p.
- Mooney, P., Corcoran, P., 2012a. The Annotation Process in OpenStreetMap. *Transactions in GIS*, 16(4), pp. 561-579.
- Mooney, P., Corcoran, P., 2012b. Who are the contributors to OpenStreetMap and what do they do? *In: Proceedings of the GIS Research UK 20th Annual Conference*, Lancaster (GBR), 11-13 April, pp. 355-360.
- Mooney, P., Corcoran, P., Winstanley, A.C., 2010. A study of data representation of natural features in OpenStreetMap, *In: Proceedings of GIScience*, Zurich (CHE), 14-17 September, pp. 150-156.
- Mooney, P., Sun, H., Yan, L., 2011. VGI as a dynamically updating data source in location-based services in urban environments, *In: 2nd International Workshop on Ubiquitous Crowdsourcing, UbiCrowd 11*, Beijing (CHN), 17-21 September 2011, pp. 13-16.
- Napolitano, M., Mooney, P., 2012. MVP OSM: A Tool to identify Areas of High Quality Contributor Activity in OpenStreetMap. *The Bulletin of the Society of Cartographers*, [In Preparation].
- Neis, P., Zielstra, D., Zipf, A., 2011. The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007-2011. *Future Internet*, 4(1), pp. 1-21.
- Neis, P., Zipf, A., 2012. Analyzing the Contributor Activity of a Volunteered Geographic Information Project—The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 1(2), pp. 146-165.
- Nielsen, J., 2006. Participation inequality: Encouraging more users to contribute, http://www.useit.com/alertbox/participation_inequality.html (Accessed 10/18, 2012).
- OpenStreetMap Wiki, 2013. API v0.6, http://wiki.openstreetmap.org/w/index.php?title=API_v0.6&oldid=890273 (Accessed 04/02, 2013).
- OpenStreetMap Wiki, 2012. Map Features, http://wiki.openstreetmap.org/w/index.php?title=Map_Features&oldid=836071 (Accessed 03/11, 2013).
- Raymond, E., 1999. The cathedral and the bazaar. *Knowledge, Technology & Policy*, 12(3), pp. 23-49.
- Rehrl, K., Gröchenig, S., Hochmair, H., Leitinger, S., Steinmann, R., Wagner, A., 2013. A conceptual model for analyzing contribution patterns in the context of VGI, *Progress in Location-Based Services, Lecture Notes in Geoinformation and Cartography*, Krisp, J.M. (Ed.), Springer-Verlag, Berlin (DEU), pp. 373-388.
- Roche, S., 2012. Should VGI map space or places? GIScience 2012 Workshop Presentation. http://www.ornl.gov/sci/gist/workshops/2012/vgi_documents/Roche.pdf (Accessed 12/18, 2012).
- van Exel, M., Dias, E., 2011. Towards a methodology for trust stratification in VGI, *In: VGI Preconference meeting at American Association of Geographers*, Seattle (USA), 3 April, 4 p.
- van Exel, M., Dias, E., Fruijtjer, S., 2010. The impact of crowdsourcing on spatial data quality indicators, *In: Proceedings of the Sixth International Conference on Geographic Information Science (GIScience 2010)*, Zurich (CHE), 14-17 September, 4 p.
- Wyllis, R.E., 1981. Empirical and theoretical bases of Zipf's law. *Library Trends*, 30(1), pp. 53-64.
- Zielstra, D., Zipf, A., 2010. A comparative study of proprietary geodata and volunteered geographic information for Germany, *In: 13th AGILE International Conference on Geographic Information Science*, Guimarães (PRT), 4 Jun 2010, 15 p.