Analysis of OpenStreetMap data quality at different stages of a participatory mapping process: Evidence from informal urban settings

Godwin Yeboah[1,*], Rafael Troilo[2], Vangelis Pitidis[1] and João Porto de Albuquerque[1]

[1] Institute for Global Sustainable Development, University of Warwick, Coventry, United Kingdom; g.yeboah@warwick.ac.uk, v.pitidis@warwick.ac.uk, j.porto@warwick.ac.uk
[2] GIScience Research Group, Heidelberg University, Heidelberg, Germany; rafael.troilo@uni-heidelberg.de

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2019 Conference in Heidelberg after peer-review.

Globally, the lack of detailed quality spatial data of informal urban settings, such as slums, is increasingly becoming a concern to both researchers and development agencies [1, 2]. One potential for making spatial data available is through Volunteered Geographic Information which is opening up new possibilities of data production in recent years and facilitating the emergence of several initiatives aimed at “putting the most vulnerable people on the map” [3]. The increasing availability of volunteered and crowdsourced geographic information, in particular OpenStreetMap (OSM), has led to a plethora of scientific studies with emphasis on evaluating the quality of the OSM data. The quality assessment results are usually presented in the form of tables, diagrams, maps and statistics per given area [4, 5]. Some recent studies have examined OSM data quality without using any external data; the so called intrinsic approach [4]. In contrast to intrinsic approach, other studies commonly used what is referred to as the extrinsic approach where the OSM data is compared with external datasets such as the UK Ordnance Survey data [6]. In both approaches, the data production processes are often not completely transparent to researchers therefore limiting possibilities for systematic data quality analysis of the processes leading to OSM update.

This presentation examines OSM data quality at different stages of a participatory mapping process developed as part of an ongoing multi-country research project focused on understanding inequalities in healthcare access of slum residents in the Global South [7]. The following research questions are addressed: (1) What is the level of spatial data quality one can expect at different stages of the mapping process leading to final update of the OpenStreetMap database? (2) What are the factors influencing quality? The OSM project provides the entire history containing all edits made by community of mappers. Examining the evolution of the OSM history data could lead to new insights about its quality and

Yeboah et al. (2019). Analysis of OpenStreetMap data quality at different stages of a participatory mapping process: Evidence from informal urban settings
In: Minghini, M., Grinberger, A.Y., Juhász, L., Yeboah, G., Mooney, P. (Eds.). Proceedings of the Academic Track at the State of the Map 2019, 7-8. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3386563

completeness. Due to the complexity, heterogeneity and size of the data, its analysis becomes a challenging endeavour. Therefore, our exploratory method applies a recently developed data analytics framework for spatio-temporal analysis, the ohsome platform [8], which allows us to analyse our study areas at different points in time. The historical data sets analysed are derived from the following stages in the participatory mapping process: before online mapping as stage 1, after online mapping and validation but before ground-truthing as stage 2, and, after ground-truthing as stage 3. The before-and-after estimates at each mapping stage are discussed together with lessons learnt. We thus present initial results from a spatial data quality assessment of the mapping process stages used to map our study areas and update the OSM database. We define completeness differential as the difference in estimates of OSM features at timestamps K and X during the participatory mapping process, where K is the feature estimate at the end of stage 2 and X is the feature estimate at the end of stage 3 or in percentage terms X divided by K times hundred.

Findings show 82% for Azam Basti (Pakistan), 99% for Sasa (Nigeria) and 63% for Korogocho (Kenya). Completeness differentials of buildings show varying differences with stage 3 showing relatively lower estimates than stage 2 for areas with unique rooftop architecture and dense structures. This means that using OSM data after stage 2 in such areas is not desirable; especially for actions such as survey research. However, for areas like Sasa with less dense structures, completeness differential is small and therefore data from stage 2 could be used for both survey research and disaster response operations with less risk. Rooftop architecture has an impact on quality of building completeness, particularly in the case of Azam Basti where the rooftop architecture representation in the satellite imagery made the interpretation of the building footprints difficult. This experience rendered stage 2 data for Azam Basti less useful during field work. Our participatory mapping process is reproducible given that the same mapping workflow and open source technologies were used across all the study sites with different local teams. However, the process still requires technical know-how and future work should focus on optimizing the integration of the tools used to make it easier to implement for survey research.

References

[1] Hachmann, S., Jokar Arsanjani, J., & Vaz, E. (2018). Spatial data for slum upgrading: Volunteered Geographic Information and the role of citizen science. *Habitat International*, *72*, 18-26.

[2] Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from space—15 years of slum mapping using remote sensing. *Remote Sensing*, *8*(6), 455.

[3] Missing Maps (2019). Putting the World's Vulnerable People on the Map. Retrieved from http://www.missingmaps.org.

[4] Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, *18*(6), 877-895.

[5] Sehra, S., Singh, J., & Rai, H. (2017). Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. *Future Internet*, *9*(2), 15.

[6] Haklay, M. (2010). How good is volunteered geographical information? A comparative analysis of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design, 37*(4), 682-703.

[7] National Institute of Health Research (NIHR) Global Health Research Unit on Improving Health in Slums project (2017). Retrieved from https://warwick.ac.uk/ fac/sci/med/about/centres/cahrd/slums

[8] Heidelberg Institute for Geoinformation Technology (2019). Big Spatial Data Analytics. Retrieved from https://heigit.org/big-spatial-data-analytics-en