

# Towards a framework for measuring local data contribution in OpenStreetMap

Maxwell Owusu<sup>1,\*</sup>, Benjamin Herfort<sup>1</sup> and Sven Lautenbach<sup>1</sup>

<sup>1</sup> Heidelberg Institute for Geoinformation Technology, Heidelberg, Germany

[maxwell.owusu@uni-heidelberg.de](mailto:maxwell.owusu@uni-heidelberg.de), [benjamin.herfort@heigit.org](mailto:benjamin.herfort@heigit.org), [sven.lautenbach@uni-heidelberg.de](mailto:sven.lautenbach@uni-heidelberg.de)

\* Author to whom correspondence should be addressed.

This abstract was accepted to the Academic Track of the State of the Map 2021 Conference after peer-review.

OpenStreetMap (OSM) has proven to be a valuable source of spatial data for many applications, including humanitarian aid. Information on buildings and roads—that can be provided by remote mapping—is of highest concern for many humanitarian applications. However, further information that can only be mapped on the ground is of high importance for finer scale humanitarian action. Road surface information, type of material and information on the use of a building (health site, school, etc.) is highly relevant. OSM offers several possibilities of adding local knowledge [1]. Recent works deal with analyzing and classifying data production in OSM [2] and intrinsic analysis has gained popularity as an indicator for measuring quality of OSM data [3–6]. Nevertheless, relatively few scientific studies have touched on "local knowledge" and local data in OSM in sufficient detail.

The question of how much local knowledge is added and what kind of local data is added remains unanswered. Addressing this question is important since only local knowledge provides access to the plethora of contextual information that is necessary for many purposes. The term "local knowledge" is often debated in the OSM community due to its ambiguity. Consequently, it is hardly taken into account by researchers when evaluating OSM [1]. This study presents a metric to measure local data contributions in OSM and analyzes temporal patterns of local contributions at three case studies. The aim of the metric is to identify archetypes of places representing a variety of contextual information.

First, we evaluated Rebecca Firth's framework on OSM contribution types that focused on the humanitarian context—see the Twitter post at [7]. Second, we discussed with local community working groups how to measure local data contributions ("What exactly are local OSM data to you?"). The outcome of the community discussion provided valuable information to design a generalized workflow for measuring local data contribution in OSM. Subsequently, we identified aspects on which the local communities agreed with respect to perception of local data. Based on these first insights, we developed a classification schema for measuring local data in OSM that is "fit-for-purpose" for local OSM communities. This schema consists of four main levels and assigned OSM tags that could be used as indicators for each level. Third, we explored the temporal evolution of local data in OSM for three unique regions. These region's mapping activities are influenced by local mapping organizations. (i) Ramani Huria in Dar es Salaam, Tanzania, focusing on flood resilience (ii)

Owusu, M., Herfort, B. & Lautenbach, S. (2021). Towards a framework for measuring local data contribution in OpenStreetMap In: Minghini, M., Ludwing, C., Anderson, J., Mooney, P., Grinberger, A.Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2021 Online Conference, July 09-11 2021, 16-18. Available at <https://zenodo.org/communities/sotm-2021>  
DOI: [10.5281/zenodo.5112234](https://doi.org/10.5281/zenodo.5112234)



Crowd2Map mainly operating in the Mara region, Tanzania and focusing on identifying features that can support the fight against girls and women at risk of female genital mutilation, and (iii) power mapping project by YouthMappers in the Koinadugu District, Sierra Leone, focusing on mapping electrical grid infrastructure. We used the osm API to access the full history of OSM. We determined the density and the ratio (as the sum of all OSM tags to the number of OSM elements) per month for each region and localness level.

The outcome of the community discussion showed that local mappers/editors had different perceptions about local knowledge. The type of local data produced depends on: (1) the context within which the data is produced and (2) the character/interest of the individual performing the mapping. However, the local data produced could be broadly categorized as "core" or "specific". The "core" category consisted of the objects that cut across almost all projects or activities (e.g., buildings, roads, place names and administrative boundary) and the category "specific" were special elements mapped as a results of a particular interest or aim of the project (e.g., culvert, drains, access types, parking type).

The developed metric is illustrated in Figure 1, showing the four main levels. Level 1 consists of objects that can be derived easily by remote mapping from satellite images such as roads and building (this is information that does not require local knowledge), level 2 focuses on place names and administrative boundaries which are frequently imported, level 3 focuses on the presence of general (e.g., residential and commercial) or specific amenities (e.g., school, clinic, and point of interest) and level 4 focuses on micro-data that provides further contextual information about an object (e.g., road:maxspeed, surface condition). Level 1 and 2 mainly fall into the "core" category whereas level 3 and 4 mainly belong to the "specific" category (which will vary across different regions).

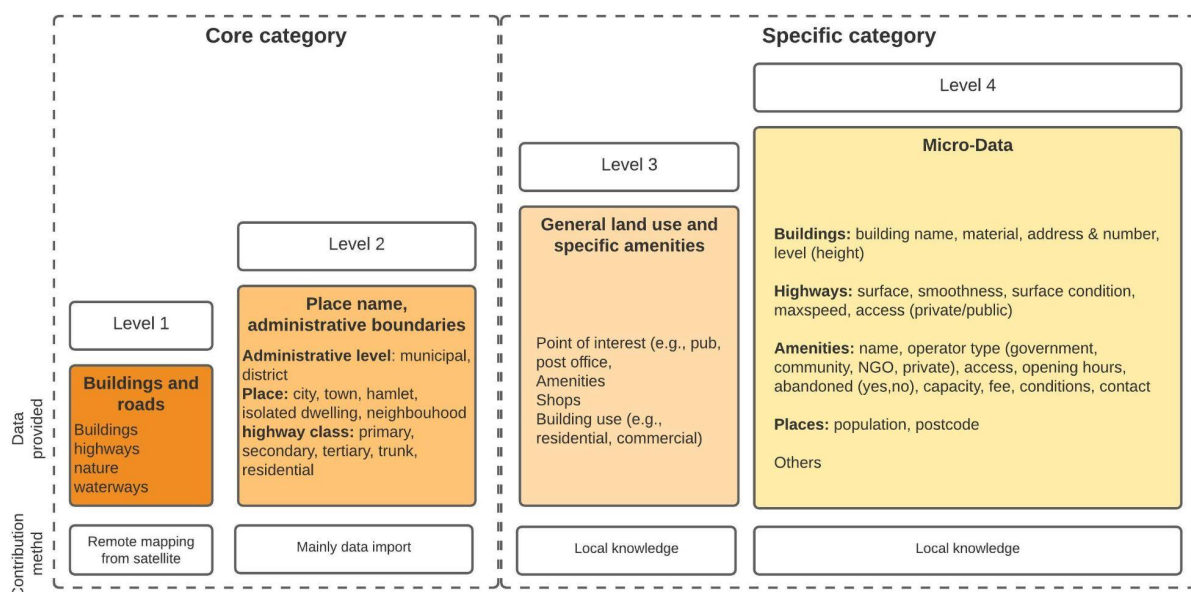


Figure 1. Classification schema for measuring local OpenStreetMap contribution.

From the temporal analysis, we observed that the amount of features in OSM decreased from level 1 to level 4. The ratio between level 1 and level 4 shows how widely local information is present in OSM at a specific location. Thereby, it provides insight on the quality of the OSM data and fitness-for-purpose for applications that need information beyond the existence of highways or buildings. Most of the mapping in the selected region

started in 2015. By digging deeper into the objects mapped, each selected region depicts unique characteristics which are largely shaped by the interest of contributors/organizations. Mapping patterns are clearly distinct from each region with respect to the development of tags. For example, there was a high amount of local data regarding waterways, drainage, and solid waste in Dar es Salaam and very low in Mara region and the Koinadugu District. It reveals the distinct mapping stories of individuals or organizations. Our results show further that there is no common path from level 1 to Level 2 to level 3 among the different regions. For the case of Dar es Salaam, mapping of features of the three levels has happened more or less simultaneously. Mapping in the Mara region focused first on place names (level 2) and then on amenities (level 3) as well as buildings and roads (level 1). For the Koinadugu District, mapping of level 2 started in 2011 and was followed by mapping of buildings and roads (level 1) in 2014 and amenities (level 3) from 2017 onward.

The classification schema helps to conceptualize a metric to measure localness of OSM data at different levels of details. This metric can be easily used to group OSM data into the categories "core" and "specific". By analyzing the temporal patterns, we identified that the contribution of local data was highly unequal and largely depended on the interest of the mapper(s). The research shed light on the richness of contextual information in OSM as well as an indication for the quality of data. In future research we would like to extend the results presented here by including more regions and more perspectives from local OSM communities. By doing so, we hope to be able to extend the definition of local data by considering the editors' local knowledge as well.

## References

- [1] Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31(1), 139-167.
- [2] Grinberger, A. Y., Schott, M., Raifer, M., & Zipf, A. (2021). An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. *Transactions in GIS*, 25, 622-641.
- [3] Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., & Zipf, A. (2021). The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific Reports*, 11(1), 1-15.
- [4] Barron, C., Neis, P., & Zipf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, 18(6), 877-895.
- [5] Klonner, C., Hartmann, M., Dischl, R., Djami, L., Anderson, L., Raifer, M., Lima-Silva, F., Degrossi, L.C., Zipf, A., & Porto de Albuquerque, J. (2021). The Sketch Map Tool Facilitates the Assessment of OpenStreetMap Data for Participatory Mapping. *ISPRS International Journal of Geo-Information*, 10(3), 130.
- [6] 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), 146-165.
- [7] Firth, R. (2019). Twitter post. Retrieved from <https://t.co/rDaSraivZE>