7-2020

Proceedings of the Academic Track at the State of the Map 2020

Marco Minghini

Serena Coetzee

Levente Juhasz

A. Yair Grinberger

Peter Mooney

See next page for additional authors

Authors
Marco Minghini, Serena Coetzee, Levente Juhasz, A. Yair Grinberger, Peter Mooney, and Godwin Yeboah

State of the Map 2020

Proceedings of the Academic Track

Editors:

Marco Minghini

Serena Coetzee

Levente Juhász

Godwin Yeboah

Peter Mooney

A. Yair Grinberger

DOI: 10.5281/zenodo.3928675

Contents

Editorial: OpenStreetMap research in the COVID-19 era

Marco Minghini[1,*,§], Serena Coetzee[2], A. Yair Grinberger[3], Godwin Yeboah[4], Levente Juhász[5] and Peter Mooney[6]

[1] European Commission, Joint Research Centre (JRC), Ispra, Italy; marco.minghini@ec.europa.eu
[2] Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa; serena.coetzee@up.ac.za,
[3] Department of Geography, The Hebrew University of Jerusalem, Jerusalem, Israel; yair.grinberger@mail.huji.ac.il
[4] Institute for Global Sustainable Development, School of Cross-faculty Studies, University of Warwick, Coventry, United Kingdom; g.yeboah@warwick.ac.uk
[5] GIS Center, Florida International University, Miami, FL, United States; ljuhasz@fiu.edu
[6] Department of Computer Science, Maynooth University, Maynooth, Ireland; peter.mooney@mu.ie

* Author to whom correspondence should be addressed.
§ The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

2020 will be remembered for a long time as the year of the coronavirus pandemic (COVID-19) caused by the SARS-CoV-2 virus [1], which so far (end of June 2020) has infected almost 10 million people and killed about 500 thousand people worldwide [2]. The pandemic, which is still ongoing around the world, has deeply affected our everyday lives and forced us to change, amongst others, the way we work and communicate. In this context, switching from physical to online conferences, as in the case of State of the Map 2020, has become an established practice. In parallel, digital technologies have shown their potential to assist societies in addressing this unexpected emergency in a variety of ways, not only limited to understanding the epidemiology of COVID-19, but also extended to the delivery of other health and public services, provision of livelihood opportunities and assurance of business continuity. Geospatial information clearly has a central role in all these processes. Its use in the COVID-19 era – powered by disruptive technologies like artificial intelligence, big data analytics and telecommunication networks – has on the one hand highlighted the potential to improve human life, but on the other hand has brought a number of issues into the spotlight, e.g. the protection of personal data, the need to still bridge the digital divide and the evidence of how the overabundance of information may contribute to the 'infodemic' [3].

There is evidence that OpenStreetMap (OSM) has been used in various ways related to the global COVID-19 pandemic during the first few months, however, exactly how still needs to be thoroughly reviewed and analysed. As always, OSM data has been contributed and used by a multitude of actors for a variety of applications [4]. The contribution of new data in OSM to address COVID-19 followed different patterns which offer fruitful ground for future research. While the Humanitarian OpenStreetMap Team (HOT) Tasking Manager (TM)

Minghini, M., Coetzee, S., Grinberger, A. Y., Yeboah, G., Juhász, L., & Mooney, P. (2020). Editorial: OpenStreetMap research in the COVID-19 era
In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2020 Online Conference, July 04-05 2020. Available at https://zenodo.org/communities/sotm-2020
DOI: 10.5281/zenodo.3922054

currently lists more than 150 projects aimed at adding baseline data, such as roads, buildings and placenames in some South-American, African and Asian countries [5], a number of specific initiatives emerged (primarily in Europe) to add or import into OSM highly localized data, such as opening hours of pharmacies or delivery/takeaway services for shops during the pandemic. The use of OSM data during the COVID-19 crisis also deserves attention from the research community. Many of the web-based dashboards produced to quantify the spread and the evolution of the pandemic use an OSM map as background, while specific OSM objects have been exploited by governments, companies, local communities and a number of other organizations to fight the pandemic, e.g. the location of pharmacies and health sites to know where people infected with COVID-19 can be treated, the road network to perform routing and the locations of pit latrines in informal settlements to map locations at risk of COVID-19 transmissions.

State of the Map 2020 features the third edition of the Academic Track, dedicated to scientific and research applications of OSM. The ten abstracts included in these proceedings have been selected by the Academic Track Scientific Committee – formed by the authors of this editorial – among all those submitted to the Academic Track call for abstracts. These ten abstracts do not directly address COVID-19-related topics as they were submitted only a few weeks after COVID-19 had been declared a pandemic. However, they cover a range of topics that are very relevant to the pandemic and at the same time provide an impression of current and future trends in OSM research.

An understanding of the quality of OSM data is essential for assessing how the data can be used. Several abstracts in this proceedings focus on methods to assess OSM quality. The first two propose a method for assessing building completeness that does not rely on local reference data, which is often not available, or ground truthing, which is expensive, and both rely on cloud computing. Biljecki and Ang [6] compare building density derived from imagery and indicators, such as the normalised difference built-up index (NDBI) and the normalised difference vegetation index (NDVI), to building density in OSM. The research aims at understanding potential coverage of 3D city models across the world. Orden et al. [7] compare OSM building completeness to Facebook's High Resolution Settlement Layer. Such quality assessment methods with global coverage can contribute to understanding the availability of geospatial data in response to a global crisis, such as the pandemic.

Looking into the future, the third abstract on quality contributes to the application of machine learning to geospatial data, e.g. for intrinsic quality assessment. Training samples and semantic information are required for such applications and are currently lacking. OSM is a freely accessible source of massive geospatial data for which comprehensive lineage information is available. Wu et al. [8] propose a flexible framework for labeling customized geospatial objects based on OSM historical data. This can be used to generate training samples and apply machine learning technologies for intrinsic assessments of data quality.

Madubedube et al. [9] employed an intrinsic quality assessment method by analysing OSM contributors and their contributions in Mozambique, a country in Southern Africa. Very few studies of OSM data on the African continent have been conducted. Similar to the first two abstracts [6, 7], the method adopted for intrinsic quality assessment is an alternative to ground truthing and to comparing OSM to local reference data. While this method does not allow absolute statements about OSM data quality, the results provide insight into the quality and can inform efforts to further improve the quality. The abstract once again demonstrates the benefits of contributing OSM data remotely, e.g. when local travel is not possible due to physical destruction of the infrastructure or prohibited to avoid risks during an epidemic or pandemic. Related to this, Schröder-Bergen [10] turns to evaluate the "localness" of OSM data as a measure of data quality, analyzing separately and on a global scale mappers' tendency to contribute within their local environment and the amount of local knowledge captured in OSM, as well as the correlation between the two.

Herforth et al. [11] present the results of an analysis of almost ten years of mapping contributions by the humanitarian mapping community using the HOT TM. These results show the impact of major mapping disaster response initiatives on OSM contributions, and how the scope widened to include other initiatives, such as disaster preparedness. These findings can inform future strategies for engaging OSM mappers. Continuing on the topic of analysing OSM contributions and contributors, Anderson and Sarkar [12] focus instead on the role of corporations in adding map data. They present a thorough analysis of the activity of three corporations within OSM – Grab, Digital Egypt, and Tesla – each producing a unique effect in terms of data contributions and the addition of new mappers to the community, in accordance with the corporations' focus on specific geographic regions and semantic themes. The OSM ecosystem developed and expanded, mainly through contributions from humanitarian and community or research initiatives. This abstract shows that the interest in contributing to OSM also originates from other reasons and that the OSM data ecosystem is suitable to address many other needs and purposes, including business-related ones.

Finally, a set of contributions to the Academic Track focus on OSM data in informal settlements or slum areas, highly relevant during the COVID-19 pandemic, as these areas are often at high risk of community infections and without proper disaster preparedness and response in place. Yeboah et al. [13] perform an OSM-based data collection for assessing spatial access to health care facilities. In the absence of detailed quality geospatial data, their approach combines field validated OSM data with the results of a health care facility survey for Sasa, an informal urban slum area in Nigeria. Results contribute knowledge on spatial proximity to health care facilities in slums through the combination of participatory remote mapping, local mapping and survey data for calculating the relations between proximity measures based on Euclidean distance and those based on network-distance. In an effort to increase the resilience of local communities to floods, Chinguwo and Mphalo [14] report the results of training local stakeholders in Malawi to produce exposure maps through the use of GPS, JOSM and secondary exposure data, showing that little work is done on the ground for mitigating flood-related hazards. This type of work could be equally useful for producing COVID-19 risk maps. Finally, Soman et al. [15] use OSM data to assess the level of spatial accessibility in 120 Low and Middle Income Countries (LIMCs) in the Global South. Based on the results, they developed a global index of under-serviced neighborhoods, which identifies the world's most spatially inaccessible communities, enabling prioritisation of infrastructure investment and, in the case of the COVID-19 pandemic, health support.

Acknowledgements

The Scientific Committee is grateful to the State of the Map 2020 Organizing Team, and particularly to Christine Karch, for the continuous support offered to the Academic Track. The Scientific Committee would also like to thank all the authors of the abstracts published in the Academic Track proceedings for contributing to the conference and for advancing OSM-related research.

References

[1] Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature medicine*, *26*(4), 450-452.
[2] World Health Organization (2020). WHO Coronavirus Disease (COVID-19) Dashboard. Retrieved from https://covid19.who.int.
[3] Mooney, P., & Juhász, L. (2020). Mapping COVID-19: How web-based maps contribute to the infodemic. *Dialogues in Human Geography*.
[4] Mooney, P., & Minghini, M. (2017). A review of OpenStreetMap data. In: Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C. C., & Antoniou, V. (Eds.) *Mapping and the Citizen Sensor*, 37-59. Ubiquity Press, London.
[5] Humanitarian OpenStreetMap Team (2020). Tasking Manager projects on COVID19. Retrieved from https://tasks.hotosm.org/explore?text=covid.
[6] Biljecki, F., & Ang, L. M. (2020). Assessing Global OpenStreetMap building completeness to generate large-scale 3D city models. In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*, July 4-5, 2020.
[7] Orden, A., Flores, R. A., Faustino, P., & Samson, M. S. (2020). Measuring OpenStreetMap building footprint completeness using human settlement layers. In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*, July 4-5, 2020.
[8] Wu, Z., Li, H., & Zipf, A. (2020). From Historical OpenStreetMap data to customized training samples for geospatial machine learning. In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*, July 4-5, 2020.
[9] Madubedube, A., Coetzee, S., & Rautenbach, V. (2020). Towards understanding the quality of OpenStreetMap contributions: Results of an intrinsic quality assessment of data for Mozambique. In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*, July 4-5, 2020.
[10] Schröder-Bergen, S. (2020). Analyzing the localness of OSM data. In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*, July 4-5, 2020.
[11] Herforth, B., Raifer, M., Stier, J., Klerings, A., & Reinmuth, M. (2020). Evolution of humanitarian mapping within the OpenStreetMap Community. In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*, July 4-5, 2020.
[12] Anderson, J., & Sarkar, D. (2020). Curious Cases of Corporations in OpenStreetMap. In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*, July 4-5, 2020.
[13] Yeboah, G., Albuquerque, J. P., & Taiwo, O. J. (2020). Examining spatial proximity to health care facilities in an informal urban setting. In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*, July 4-5, 2020.
[14] Chinguwo, D., & Mphalo, B. (2020). Community mapping as a means to building resilience. In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*, July 4-5, 2020.
[15] Soman, S., Nederhood, C., Marchio, N., Yang, A., Beukes, A., & Bettencourt, L. (2020). Detecting informal settlements via topological analysis. In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*, July 4-5, 2020.

Assessing global OpenStreetMap building completeness to generate large-scale 3D city models

Filip Biljecki[1,]* and Li Min Ang[1]

[1] Urban Analytics Lab, National University of Singapore, Singapore; filip@nus.edu.sg, limin.ang1@gmail.com

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2020 Online Conference after peer-review.

Quality assessment of OpenStreetMap (OSM) data has been an important topic since the inception of the project [1]. Much research has been done on this topic by many research groups around the world, and it can mostly be seen as permutations of three aspects: (1) spatial data quality element(s) in focus (e.g. positional accuracy, completeness), (2) theme (e.g. amenities, buildings, roads), (3) geographical area (e.g. particular city or country); e.g. positional accuracy of cultural features in Italy [2].

Completeness is one of the principal quality aspects of geospatial data, and our research focuses on developing a method to assess the completeness of buildings in OSM on a large scale (spanning several countries). While there are many robust OpenStreetMap completeness techniques and studies developed, they mainly focus on limited areas, mostly developed countries with ground truth data at hand for comparisons.

Doing the same for less developed regions is rare as the lack of authoritative data inherently hampers it, and the methods hardly ever scale: such an analysis done simultaneously for more than one administrative region is seldom carried out because there are other research challenges such as disparate urban morphology, different data sources and standards to bridge in order to facilitate ground truth, and varying understanding of what a building is. Furthermore, the development of a method that would scale across dozens of countries is limited by computational resources.

We are currently developing a method that uses several indicators derived from remote sensing, which are available on the global level, that may hint at the building completeness and would scale across the world. A regression model to predict the approximate volume of buildings in a given area is trained in areas in which there is an indication of high completeness of buildings in OSM.

OSM building completeness is estimated by comparing the area of mapped buildings against their expected (predicted) coverage in reality. The method has the potential to scale worldwide, and completeness is estimated for a grid of resolution of approximately 1x1 square kilometres, and simultaneously for administrative regions to enable cross-country comparisons.

Biljecki, F., & Ang, L. M. (2020). Assessing Global OpenStreetMap building completeness to generate large-scale 3D city models In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2020 Online Conference, July 4-5 2020. Available at https://zenodo.org/communities/sotm-2020
DOI: 10.5281/zenodo.3922285

The work is being implemented in Google Earth Engine, mostly relying on imagery and indicators such as normalised difference built-up index (NDBI) and normalised difference vegetation index (NDVI). Preliminary results suggest a substantial disparity in OSM building completeness around the world, with areas that are entirely complete to those with inadequate completeness.

This preliminary paper aims to report the progress and the encountered challenges in the ongoing project such as selecting indicators that are consistently available on the global level, applying the method widely, accounting for different mapping practices around the world, different notions of a building, and varying morphologies of cities and urban areas. We also investigate the relation between OSM building completeness and socio-economic parameters such as the gross domestic product (GDP) to understand their relationships to mapping intensity and quality. These may offer the potential to be used as additional predictors.

This work is part of a broader project conducted at the Urban Analytics Lab at the National University of Singapore on investigating the potential of generating 3D city models by extruding building footprints in OpenStreetMap to a building height that is predicted using artificial intelligence [3]. This ongoing portion of the project is the crucial first step in the project, as it will enable us to understand what is the completeness of building footprints in OpenStreetMap around the world and manage expectations about the potential coverage of 3D city models.

Acknowledgements

This research is part of the project Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start Up Grant R-295-000-171-133.

References

[1] Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31(1), 139-167.

[2] Balducci, F. (2019). Is OpenStreetMap a good source of information for cultural statistics? The case of Italian museums. *Environment and Planning B: Urban Analytics and City Science*.

[3] Biljecki, F., Ledoux, H., & Stoter, J. (2017). Generating 3D city models without elevation data. *Computers, Environment and Urban Systems*, 64, 1-18.

Measuring OpenStreetMap building footprint completeness using human settlement layers

Ardie Orden[1,*], Ren Avell Flores[1], Pia Faustino[1] and Mark Steve Samson[1]

[1] Thinking Machines Data Science, Taguig, Philippines; ardie@thinkingmachin.es, avell@thinkingmachin.es, pia@thinkingmachin.es, marksteve@thinkingmachin.es

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2020 Online Conference after peer-review.

Non-government organizations and local government units use geographic data from OpenStreetMap (OSM) to target humanitarian aid and public services. As more people start to depend on OSM, it is important to study data completeness in order to identify unmapped regions so that OSM volunteers can focus their attention on these areas. In this study, we propose a method to measure the data completeness of OSM building footprints using human settlements data.

Specifically, we use Facebook's High Resolution Settlement Layer (HRSL), a dataset of built-up areas derived from satellite images, as a proxy for ground truth building footprints. We then measure data completeness by getting the "percentage completeness" of pixels which is computed using the total percentage of pixels within the intersection of the human settlement layer and the OSM building footprints. The method can be broken down into three steps: (1) convert the human settlement layer into a vector; (2) perform a spatial join to find the intersection between the vectorized human settlement layer and the OSM building footprints; and (3) calculate the data completeness based on areas from the vectorized human settlement layer that intersect with the building footprints.

Chepeish and Polchlopek [1] conducted a similar study measuring data completeness in OSM building footprints. We differentiate our work from Chepeish and Polchlopek [1] in three ways. First, for the human settlement layers, they used WorldPop which has a spatial resolution of 100 meters [2] while we used the High Resolution Settlement Layer (HRSL) from Facebook which has a spatial resolution of 30 meters [3]. Second, for the data processing, they rasterized the building footprints while we vectorized the human settlement layer. Third, for calculating the data completeness, they used a combination of geographic information system (GIS) and machine learning (ML) while we solely used GIS.

Using building footprints from January 2020 and human settlement layers dated June 2019 and October 2018, the percentage completeness is 32.75% and 10.89% for the Philippines and Madagascar, respectively. We found that in the Philippines, most of the unmapped buildings are in rural areas. When the pixels are aggregated to the municipality-level and plotted as a scatter plot of the urban percentage completeness vs. the rural percentage completeness, the municipalities appear to group together into two

Orden, A., Flores, R. A., Faustino, P., & Samson, M. S. (2020). Measuring OpenStreetMap building footprint completeness using human settlement layers
In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., Grinberger, & A. Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2020 Online Conference, July 4-5 2020. Available at https://zenodo.org/communities/sotm-2020
DOI: 10.5281/zenodo.3923033

categories: sparsely mapped and thoroughly mapped. A possible explanation is that there are not enough OSM volunteers to map all municipalities and the OSM community focuses on thoroughly mapping high population municipalities rather than moderately mapping all municipalities. Interestingly, poverty incidence data from the Philippine Statistics Authority is not correlated with data completeness. Complete or incomplete OSM data in an area is not an indicator of wealth or poverty.

As this work has garnered interest from humanitarian organizations such as the Humanitarian OpenStreetMap Team (HOT), to whom regularly updated information on OSM data completeness is extremely valuable, we looked into ways to automate workflows in QGIS by using the built-in workflow builder tool (i.e. Processing Modeller) and by using the QGIS API. However, as we consider scalability and reproducibility important for this line of work, we ultimately deemed QGIS to be unfit for our use case. QGIS is not scalable because the data processing is not easily parallelizable and it is also not easily reproducible by developers who do not have training with GIS software. Thus, we decided to migrate our workflow to GeoPandas and rasterio and open-source our code [4]. Our workflow improved because (1) we were able to speed up the process by migrating to the cloud and increasing the computing resources; and (2) we were able to improve the reproducibility by allowing us to communicate our work more effectively to people who are not familiar with GIS.

For future research, we recommend exploring other human settlement datasets. The Global Human Settlement Layer (GHSL), for example, has a spatial resolution of 30 meters [5] which is comparable to WorldPop and HRSL. We also encourage further data analysis on the percentage completeness in order to get insights on how to improve the process of contributing to OSM.

References

[1] Chepeish, E., & Polchlopek, J. (2018). Estimate OSM building coverage completeness by comparing vs WorldPop raster. Retrieved from https://github.com/azavea/hot-osm-population.

[2] Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific data*, *4*(1), 1-4.

[3] Tiecke, T. G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., Prydz, E. B., & Dang, H. H. (2017). Mapping the world population one building at a time. *arXiv preprint arXiv:1712.05839*.

[4] Orden, A., & Flores, R. A. (2020). Supplementary code for "Measuring OpenStreetMap building footprint completeness using human settlement layers. Retrieved from https://github.com/thinkingmachines/osm-completeness.

[5] Florczyk, A. J., Corbane, C., Ehrlich, D., Freire, S., Kemper, T., Maffenini, L., Melchiorri, M., Pesaresi, M., Politis, P., Schiavina, M., Sabo, F., & Zanchetta, L. (2019). GHSL Data Package 2019. JRC Technical Report, EUR 29788 EN, Publications Office of the European Union, Luxembourg.

From Historical OpenStreetMap data to customized training samples for geospatial machine learning

Zhaoyan Wu[1,2], Hao Li[1,*] and Alexander Zipf[1]

[1] GIScience Research Group, Heidelberg University, Heidelberg, Germany;
zhaoyan.wu@uni-heidelberg.de, hao.li@uni-heidelberg.de, zipf@uni-heidelberg.de
[2] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China;

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2020 Online Conference after peer-review.

After more than a decade of rapid development of volunteered geographic information (VGI), VGI has already become one of the most important research topics in the GIScience community [1]. Almost in the meantime, we have witnessed the ever-fast growth of geospatial machine learning technologies to develop intelligent GIServices [2] and to address remote sensing tasks [3], for instance land use/land cover classification, object detection, and change detection. Nevertheless, the lack of abundant training samples as well as accurate semantic information has been long identified as a modelling bottleneck of such data-hungry machine learning applications. Correspondingly, OpenStreetMap (OSM) shows great potential in tackling this bottleneck challenge by providing massive and freely accessible geospatial training samples [4, 5]. More importantly, OSM has exclusive access to its full historical data [6], which could be further analyzed and employed to provide intrinsic data quality measurements of the training samples. Therefore, a flexible framework for labeling customized geospatial objects using historical OSM data allows more effective and efficient machine learning.

This work approaches the topic of labeling geospatial machine learning samples by providing a flexible framework to automatically generate customized training samples and provide intrinsic data quality measurements. In more detail, we explored the historical OSM data for two purposes: feature extraction and intrinsic assessment. For example, when training building detection convolutional neural networks (CNNs), the OSM features with tags as *building=residential* or *building=house* are certainly of interest while the data quality of such features might play an important role later in the CNNs training phase. Therefore, besides the acquisition of the user-defined OSM features, we provide additional intrinsic quality measurements. Currently, we consider some basic statistics, such as the areas of buildings tagged with different OSM tags, the number of distinct contributors in the last six months, or the equidistance of polygons with *landuse=cropland* etc., since the existing research suggested that the lower equidistance of the current polygon, the better the relative quality of the polygon, which is due to the further refining and editing from users [7]. In the

Wu, Z, Li, H., & Zipf, A. (2020). From Historical OpenStreetMap data to customized training samples for geospatial machine learning
In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2020 Online Conference, July 4-5 2020. Available at https://zenodo.org/communities/sotm-2020
DOI: 10.5281/zenodo.3923040

future, one could also easily extend the current framework and develop other sophisticated quality indicators for specific "fitness-for-use" purposes.

Heterogeneous remote sensing APIs are supported within the framework. User options range from commercial satellite image providers (e.g., Bing or Mapbox) to government satellite missions (e.g., Sentinel-hub) and even user-defined tile map service (TMS) APIs. Corresponding to OSM features, the satellite image would be automatically downloaded via TMS and tiled into proper size. Moreover, this framework supports different machine learning tasks, such as classification, object detection, and semantic segmentation, which requires distinct sample formats. The preliminary test is performed to extract the geographical information of water dams with OSM tag *waterway=dam,* which enables the training of water dams detection CNNs, where users could easily change the geospatial water dams to customized objects as long as the corresponding OSM tags are identified.

The aim of this work is to promote the application of geospatial machine learning by generating and assessing OSM training samples of user-specified objects, which not only allows user to train geospatial detection models, but also introduces the intrinsic quality assessment into the "black box" of the training of machine learning models. Based on a deeper understanding of training samples quality, future efforts are needed towards more understandable and geographical aware machine learning models.

References

[1] Yan, Y., Feng, C. C., Huang, W., Fan, H., Wang, Y. C., & Zipf, A. (2020). Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. *International Journal of Geographical Information Science*, 1-27.

[2] Yue, P., Baumann, P., Bugbee, K., & Jiang, L. (2015). Towards intelligent giservices. *Earth Science Informatics*, 8(3), 463-481.

[3] Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8-36.

[4] Li, H., Feld, I. N., Herfort, B., & Zipf, A. (2019). Estimating OpenStreetMap Missing Built-up Areas using Pre-trained Deep Neural Networks. *Proceedings of the 22nd AGILE Conference on Geographic Information Science*, Limassol, Cyprus, 17-20.

[5] Chen, J., & Zipf, A. (2017). DeepVGI: Deep learning with volunteered geographic information. *Proceedings of the 26th International Conference on World Wide Web Companion*, Perth, Australia, 771-772.

[6] Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., Przybill, K., Fendrich, S., Mocnik, F.-B., & Zipf, A. (2019). OSHDB: a framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospatial Data, Software and Standards*, 4(1).

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

Towards understanding the quality of OpenStreetMap contributions: Results of an intrinsic quality assessment of OpenStreetMap for Mozambique

Aphiwe Madubedube[1], Serena Coetzee[1,*] and Victoria Rautenbach[1]

[1] Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa; u12328929@tuks.co.za, serena.coetzee@up.ac.za, victoria.rautenbach@up.ac.za

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2020 Online Conference after peer-review.

OpenStreetMap (OSM) has made it possible for any volunteer to contribute geographic information, regardless of their level of experience or skills. Since the task of creating geographic information is no longer exclusively performed by trained professionals, data quality can be a concern and has been cited as a hindrance to its use [1]. OSM data quality has often been assessed extrinsically by comparing it to other reference datasets. However, such reference data is not always available and some feature definitions (e.g. wetlands) are subjective. Therefore, intrinsic assessment methods, where the data itself is analysed, have been employed. For example, analysing contributors and their contributions can answer questions, such as: What kind of contributors (e.g. experienced vs newcomers) have worked on the data in the area? In which areas should the data be validated or updated (e.g. where data has been contributed by newer contributors or older non-recurring contributors). As the number of OSM contributors continues to grow, knowledge about their characteristics and the kind of data they contribute is increasing in importance.

In this study, contributors and their contributions to OSM in Mozambique, a country in Southern Africa, were analysed. We chose Mozambique because it received a significant amount of attention in the OSM community following the floods and damages as a result of cyclones Idai and Kenneth in 2019. OSM contributors were characterised in three steps: 1) OSM history data was downloaded, containing information about contributors and their contributions; 2) using k-means cluster analysis and the elbow method for determining an optimal number of clusters, OSM contributors were classified according to the contribution timespan, intensity of editing activities and the type of editing done; 3) based on the classification, contributors were characterised to gain insight into the quality of the data.

OSM history data provides a record of all the edits (or changes) performed on OSM features. Each edit results in an increment in a feature's version number. Each version of a feature is associated with a contributor (called 'user'). The results of the cluster analysis revealed four distinct classes of contributors: most active (n=2552); older non-returning (2676); least productive (n=1301); and recent, perhaps attracted by the disasters (n=3708).

Madubedube, A., Coetzee, S., & Rautenbach, V. (2020). Towards understanding the quality of OpenStreetMap contributions: Results of an intrinsic quality assessment of data for Mozambique
In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2020 Online Conference, July 4-5 2020. Available at https://zenodo.org/communities/sotm-2020
DOI: 10.5281/zenodo.3923042

The most active class of contributors represented 25% of all contributors in the area, with on average the highest numbers of changesets, total contributions, node contributions, way contributions and ways and nodes for which they were the last user to modify them. These volunteers are 'older' contributors who have sustained their contributions in the Mozambique area over a long period of time, with first contributions dating back around 15 years. Compared to other contributors, they have mapped on more days than the others and the average number of edits per feature is higher. Nevertheless, 99.6% of buildings and 84% of ways in the data had been edited twice at most, and most of the buildings appeared in the areas for which mapathons were conducted, outside larger cities. In other parts of the world, editing is more frequent in densely populated areas of cities. More studies in different parts of Africa would contribute to understanding whether the Mozambican pattern is unique or also found in other parts of the African continent.

Similar to the results of other OSM contribution analyses [2], most of the data generated in Mozambique was contributed by a small group of active contributors who have dedicated a significant amount of time to this. Studies have suggested that such contributors are more likely to be experienced and knowledgeable about the project [3, 4] and are therefore more likely to produce data that is of good quality [3, 4]. In this study, individual contributions were clustered, it would be interesting to cluster changesets to assess whether contributor behaviour is consistent across different changesets. Visualizing OSM data based on its history could also provide a quick quality assessment [5].

Even though no absolute statements can be made about the quality of the OSM data for Mozambique, analysing contributors and their contributions provides valuable insight into the quality of the data and can inform efforts to further improve the quality. The results of this study show how one can gain a better understanding of the community that contributes data in a specific area by inspecting history data. Such intrinsic methods should not replace ground truthing or extrinsic methods, but rather complement them by providing alternative ways to gain insight about data quality.

References

[1] Mooney, P., & Morgan, L. (2015). How much do we know about the contributors to volunteered geographic information and citizen science projects? *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3/W5, 339-343.
[2] 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.
[3] Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, 18(6), 877-895.
[4] Yang, A., Fan, H., & Jing, N. (2016). Amateur or professional: Assessing the expertise of major contributors in OpenStreetMap based on contributing behaviors. *ISPRS International Journal of Geo-Information*, 5(2), 21.
[5] Minghini, M., & Frassinelli, F. (2019). OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date? *Open Geospatial Data, Software and Standards,* 4(9).

Curious Cases of Corporations in OpenStreetMap

Jennings Anderson[1,]* and Dipto Sarkar[2]

[1] Department of Computer Science, University of Colorado Boulder, Boulder, USA; jennings.anderson@colorado.edu
[2] Department of Geography, University College Cork, Cork, Ireland; dipto.sarkar@mail.mcgill.ca

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2020 Online Conference after peer-review.

OpenStreetMap (OSM), the largest crowdsourced geographic database has garnered interest from corporations over the last four years. Today, major corporations including Apple, Facebook, and Microsoft have dedicated teams contributing to OSM. More than 2,300 OSM editors are associated with corporate data teams, up from approximately 1,000 in 2019 [1]. As of March 2020, nearly 17% of the global road network (measured per kilometer) was most recently edited by a corporate data-team member. Each corporation edits according to their own agenda; displaying unique patterns of edits with respect to types of features edited, mode (manual, import, or machine assisted), locations, and volume of edits.

We investigate the unique editing patterns associated with three corporations: Grab, Digital Egypt, and Tesla, the latter two's editing activity has never previously been quantified. Differing from corporations with high volumes of global editing [1], these corporations operate in specific geographies and exhibit uniquely specific patterns. We use a combination of OSM data processing pipelines including tile-reduce, osm-qa-tiles, and osm-interaction tilesets [2] to extract and quantify the edits associated with the corporations.

Grab is a Singapore-based company active in South-East Asia offering ride-hailing transport, food delivery, and payment services. Grab is actively editing OSM data since 2018 and has thus far edited 1.6M features. Grab's focus on transport related services implies that a navigable road network is a priority. However, topology and navigation restrictions are difficult to encode. Grab dedicates efforts to improve road navigability. In Singapore, Grab has edited over 100,000 turn restrictions, comprising 95% of all turn restrictions in Singapore (and 7% of all turn restrictions globally). This represents a highly focused effort put in by a corporation in a specific place to build infrastructure needed to support their business. Overall, Grab's efforts of improving data and building a community of editors in South-East Asia is beneficial for the OSM ecosystem.

Digital Egypt (DE) aims to produce detailed GIS and Mapping data in Egypt, Middle East and Africa. Active in a part of the world with sparse geographic data coverage, Digital Egypt's team of 24 mappers is working to improve the accuracy of OSM for geocoding. As of March 2020, DE has edited more than 2M features in Egypt, more than 1.7M of these edits involve objects with address tags (e.g. *addr:housenumber*). These edits comprise 94% of the objects in Egypt that have an address tag. Similar to Grab, DE's contributions improve the

Anderson, J., & Sarkar, D. (2020). Curious Cases of Corporations in OpenStreetMap
In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2020 Online Conference, July 4-5 2020. Available at https://zenodo.org/communities/sotm-2020
DOI: 10.5281/zenodo.3923050

usability of OSM data for everyone. Unlike Grab, DE does not operate their own services (such as routing) on top of this data infrastructure. Their hyperlocal focus and dedication to improving data for a country with a dearth of spatial data distinguishes them from other corporations who are using the map data within a product by instead making the local data the product itself.

Tesla—an American manufacturer of electric cars—was revealed to be using OSM parking aisle information for their vehicle's self-driving "summon" feature in a blog post in November 2019 [3]. For proper functionality, the summon feature requires detailed maps of parking aisles within larger parking lots. These data were relatively sparse, as their utility to the overall map is minimal compared to the actual road network. However, the number of parking aisles added to OSM in North America has increased by approximately 71% in 2018-2020 from an average of 322 ways per day in 2016-2018. Subsequently, the number of daily editors adding parking aisles within their first week of joining OSM spiked in the days following the blog. While there is no official, disclosed Tesla data team that is mapping these features, in the days following the blog post, the number of new OSM editors adding parking aisles within their first week of editing jumped from an average of 1.5 to 10.

This single blog post inspired scores of Tesla fans to join OSM and add new parking aisles to the map, thereby mapping a specific feature type for a narrow purpose. These Tesla owners represent a new generation of hobby and "craft mappers" in OSM. Unlike traditional 'craft-mappers' considered to altruistically contribute map data about specific features, the Tesla mappers create data intended to be consumed by a corporation to enhance the experience for a select group of motorists. If looking to categorize these new mappers into prior community labels, their mapping practices have more in common with a "traditional craft-mapper" than a paid, corporate mapper.

Grab, Digital Egypt, and Tesla represent three distinct cases of corporations consuming, contributing, and driving further interest in OSM.

References

[1] Anderson, J., Sarkar, D., & Palen, L. (2019). Corporate editors in the evolving landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 8(5), 232.

[2] Anderson, J. (2020). osm-interactions. Retrieved from https://github.com/jenningsanderson/osm-interactions.

[3] Tesla Motors Club (2019). Tesla owners can edit maps to improve summon routes. Retrieved from https://teslamotorsclub.com/blog/2019/11/04/tesla-owners-can-edit-maps-to-improve-summon-routes.

Examining spatial proximity to health care facilities in an informal urban setting

Godwin Yeboah[1,*], João Porto de Albuquerque[1] and Olalekan John Taiwo[2]

[1] Institute for Global Sustainable Development, University of Warwick, Coventry, United Kingdom; g.yeboah@warwick.ac.uk, j.porto@warwick.ac.uk
[2] Department of Geography, University of Ibadan, Nigeria; oj.taiwo@mail.ui.edu.ng

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2020 Online Conference after peer-review.

The United Nations Sustainable Development Goal (SDG) 3 seeks to ensure universal health coverage for everyone irrespective of geographical location by 2030. Anecdotal evidence exists on the possibility of attainment of the goal at household level in slum areas most especially in Africa. Recent studies suggest that further work is needed on the advancement of empirical evidence on slum health [1, 2], especially in Africa where slum population growth is reported to be at the same level with urban population growth [3]. There is the need to understand the dimensions of spatial proximity to healthcare facilities in Nigeria towards achieving SDG 3 [4], especially in slum areas where little evidence exists. Spatial access to appropriate healthcare is even more relevant given the rapid rise in Covid-19 cases globally. The lack of detailed quality spatial data is a concern to both researchers and development agencies [5]. In an attempt to contribute to the knowledge gap in slum health studies, this study draws on two data sets (field validated OpenStreetMap data and healthcare facility survey data) from an ongoing research project to examine spatial proximity to healthcare facilities (HCFs) in Sasa, an informal urban "slum" area in Nigeria. The decision to focus on spatial proximity is based on findings from a household survey, in an ongoing project, which suggest that one of the main reasons given by respondents for choosing HCFs is proximity. Conceptually, there are two main schools of thought about spatial proximity [6]; this study considers proximity as a distance measure defined quantitatively. The ongoing research project is a National Institute for Health Research (NIHR) Global Health Unit on Improving Health in Slums at University of Warwick [7]. This Unit focuses on health services in slums through the study of seven slum sites in Africa and Asia and aims at finding optimal ways to enhance health services. We thus present initial results from one of the study sites in Africa. The following research questions are explored. What are the differentials of spatial proximity to health care providers in informal settlements like slum? What are some of the lessons learnt from using OpenStreetMap-based mapping approach for slum health research?

An OpenStreetMap-based data collection methodological approach was developed and implemented [8]. A spatially-referenced sampling frame was generated through a combination of: remote participatory mapping from satellite imagery; local participatory mapping and ground-truthing; and the identification of dwellings of each validated structure.

Yeboah, G., Albuquerque, J. P., & Taiwo, O. J. (2020). Examining spatial proximity to health care facilities in an informal urban setting
In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2020 Online Conference, July 4-5 2020. Available at https://zenodo.org/communities/sotm-2020
DOI: 10.5281/zenodo.3923053

Additionally, a healthcare facility survey was conducted to capture types of facilities etc. The following categories of HCFs are drawn from survey data and used for analyses: four Clinics and Maternity Centres (CMC); twenty-two Patent Medicine Stores (PMS); five Traditional and Faith Healers (TFH); and one Eye Health Centre (EHC). Two-fold analyses are conducted. First, two measures of spatial proximity (spatial network and Euclidean distances) to different types of HCFs within the site are computed using field validated OpenStreetMap (OSM) network data. Bivariate analysis is performed to test sinuosity (ratio of network and Euclidean distance). Additionally, comparative analyses of combined means and medians (using k-independent samples median tests) for categories of HCFs are performed. Second, a reflective exercise is undertaken to outline some of the lessons learnt during the research process related to the OSM-based approach.

This study presents the outcome of the two-fold inquiry outlined. Preliminary results show strong positive correlation (r=0.97; 99% CI) between the two spatial proximity measures suggesting that Euclidean and network spaces are quite similar in terms of accessibility to health care services within Sasa slum. Overall sinuosity index is 1.16 suggesting that the non-linear nature of network routes to HCFs contributes to 16% more than the Euclidean metric. The combined network distance grand mean (with standard deviations) and grand medians for each of the categories are as follows: 727m (±299) and 766m for CMC; 579m (±256) and 563m for PMS; 589m (±240) and 589m for TFH; and 503m (±204) and 490m for EHC. Residents can access these facilities within a walking distance (under 1km) where Clinics and Maternity Centres appear to be the farthest from most residents. This study advances the evidence base on slum health towards achieving SDG 3 and promotes the use of OSM-based mapping approach and data for slum health research.

References

[1] Ezeh, A., Oyebode, O., Satterthwaite, D., Chen, Y. F., Ndugwa, R., Sartori, J., Mberu, B., Melendez-Torres, G. J., Haregu, T., Watson, S. I., Caiaffa, W., Capon, A., & Lilford, R. J. (2017). The history, geography, and sociology of slums and the health problems of people who live in slums. *The lancet*, 389(10068), 547-558.

[2] Lilford, R., Taiwo, O. J., & de Albuquerque, J. P. (2018). Characterisation of urban spaces from space: going beyond the urban versus rural dichotomy. *The Lancet Public Health*, 3(2), e61-e62.

[3] Badmos, O. S., Callo-Concha, D., Agbola, B., Rienow, A., Badmos, B., Greve, K., & Jürgens, C. (2020). Determinants of residential location choices by slum dwellers in Lagos megacity. *Cities*, 98, 102589.

[4] Lawal, O., & Anyiam, F. E. (2019). Modelling geographic accessibility to primary health care facilities: Combining open data and geospatial analysis. *Geo-Spatial Information Science*, 22(3), 174-184.

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

[6] Brennan, J., & Martin, E. (2012). Spatial proximity is more than just a distance measure. *International journal of human-computer studies*, 70(1), 88-106.

[7] Lilford, R. (2017). NIHR Global Health Research Unit on Improving Health in Slums project at the University of Warwick. *National Institute of Health Research (NIHR)*.

[8] Improving Health in Slums Collaborative (2019). A protocol for a multi-site, spatially-referenced household survey in slum settings: methods for access, sampling frame construction, sampling, and field data collection. *BMC medical research methodology*, 19(109), 1-8.

Community mapping as a means to building resilience

Dickson Chinguwo[1,]* and Brown Mphalo[2]

[1] Computing and Information Technology Department, University of Malawi, The Malawi Polytechnic, Blantyre, Malawi; dixchinguwo@gmail.com
[2] Surveys Department, Ministry of Lands, Housing and Urban Development, Lilongwe, Malawi; mphalobrown@gmail.com

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2020 Online Conference after peer-review.

The greatest threat to the people, property and economy of Malawi is represented by natural hazards. Since 1946 of all the 298 times the country has been impacted by hazards, 89% of those have been natural hazards while the remaining 11% have been human-generated events [1]. Disaster records for the last 100 years indicate that Malawi has experienced about 20 droughts while in the last 36 years alone, the country has experienced 8 major droughts which have affected over 24 million people [2].

In Malawi, quite a number of districts are hit by floods almost every year. Households and infrastructures such as roads, schools and praying houses are affected in much the same way as cultivation fields. These disasters impact negatively on the socio-economic growth of the communities involved and the country at large to some extent. For example, MacOpiyo [2] pointed out that in 2015 Malawi experienced a once-in-500-years flood which impacted more than 1.1 million people. Therefore, this study aimed at collecting exposure data which was used for production of flood risk maps which were in turn used for flood risk Atlas production.

Primary and secondary data were used in this study. Primary data which was exposure data such as buildings, toilets, roads, bridges and schools among others was collected using handheld Global Positioning System (GPS) devices with an accuracy of 3m. To ensure precision on features captured using the handheld GPS devices and to easily identify features captured on the ground, collected coordinates of exposure data was overlaid with satellite imagery in Java OpenStreetMap (JOSM) software when digitizing, creating attributes for that data and uploading the digitized layers into OpenStreetMap. Secondary data came from the disaster profile (database) of the Department of Disaster and Management Affairs (DoDMA) in Malawi which helped in identifying flood prone areas. Choice of this data was based on how it impacts the socio-economic activities of the concerned communities. Before each mapping activity, workshops were organized in each of the 15 districts. Participants with different backgrounds (cartographers, water resource managers, academicians, surveyors, IT personnel) were trained in the use of GPS, JOSM and Geographic Information Systems (GIS) in general.

Chinguwo, D., & Mphalo, B. (2020). Community mapping as a means to building resilience
In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2020 Online Conference, July 4-5 2020. Available at https://zenodo.org/communities/sotm-2020
DOI: 10.5281/zenodo.3923058

One of the outcomes of this research was maps highlighting some statistics of the exposure data under study. These statistics show captured features on the ground and the extent to which the hazard (flood) affected the communities. Such statistics were generated by performing some analysis through InaSAFE, an extension of the QGIS software - a popular open-source GIS with advanced capabilities. The maps and the underlying analysis helped to answer questions like: in an event of floods in a particular area, how many households, people, and infrastructures will be affected?

Such statistics will help in terms of planning, recovery and even response to the floods. The maps also help along the same lines of planning, response and recovery. The maps also help communities come up with contingency plans, including evaluating the needs and cooperating quickly. Because of the maps, the communities build resilience as the maps display areas where they can settle or not. The flood hazards are displayed on the map by their spatial location. The maps also also show the impact of the floods; as a result people are made aware of the threats and hence build resilience.

From the study, it was noticed that there are little interventions on the ground to help in reducing the root cause of floods in most parts of the country. There are a number of reasons for the causes of floods in the affected areas. Siltation in rivers causes river beds to rise-up which in-turn causes floods. Living in low lying areas in search for fertile land is another element that was found from communities to be heavily affected by floods. Another problem that was observed during the course of this study is the lack of knowledge and information on disaster risk management, which has resulted in communities being affected when flood disasters strike. The study contributed towards some best practices in carrying out community mapping exercises as well as freely distributing results on OSM and spatial data portals like the Malawi Spatial Data Platform (MASDAP) for further studies and/or decision making. Thus, the study focused on preparing for mapping – what to map, how to map and how to record the data; the mapping exercise itself; the download and digitization of data in map production; and the use of maps to aid decision making. The flood risk maps and the Atlases are a form of awareness to the communities and the country at large about flood risk within their vicinity; as a result, they help build resilience among the communities.

As a way forward, the study proposes that the government and other stakeholders provide communities with long-term interventions in building their capacities and resilience to reduce vulnerabilities. This might be in the form of building of dykes along river banks to contain floods or the enforcement of proper construction standards when putting up infrastructures. Communities are also encouraged to build permanent dwelling houses with strong foundations based on standard guidelines for safer housing and construction.

References

[1] Misomali, R. (2014). Emergency Management In Malawi: A Work in Progress With a Strong Foundation.

[2] MacOpiyo, L. (2017). Community Based Resilience Analysis (COBRA) Assessment Report for Zomba, Ntcheu and Nkhata Bay Districts in Malawi. *United Nations Development Programme*.

[3] MASDAP (2020). Malawi Spatial Data Platform - A public platform for GIS Data to support development in Malawi. Retrieved from http://www.masdap.mw.

Analyzing the localness of OSM data

Susanne Schröder-Bergen[1,*]

[1] Institute of Geography, Friedrich-Alexander University Erlangen-Nürnberg, Germany; susanne.schroeder-bergen@fau.de

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2020 Online Conference after peer-review.

A regularly reproduced mantra of the debate on Volunteered Geographic Information (VGI) can be summarized by the term “localness”: it emphasizes the actual or at least the desired local production of geographical information. It is assumed that many “ordinary” people “on the ground” can now generate knowledge about their everyday environment through new tools in Web 2.0. This postulated “local expertise” operates as a claim to truth of VGI. In contrast to “conventional” geographical data, whose truth claim is more likely to be based on professional and technical expertise, VGI is often legitimized by its authenticity due to its “localness” [1, 2].

In fact, relatively little is known about how “local” VGI data actually are. Although “localness” is accepted as a central quality feature of VGI [3], it is hardly taken into account for its evaluation [4]. In contrast, the importance of “localness” is controversially discussed in the OpenStreetMap (OSM) community. Our key interest is therefore to further investigate the significance and relevance of “localness”. We started with a synopsis of all methods that have been already used to measure “localness”. A first insight leads to the observation that many of the used methods work best at regional scale, only some function for a global comparison. Secondly, an examination of the already implemented methods suggests that there are structural differences in the evaluated “local material”. On the one hand, the approaches pursue the goal of determining the geographic origin or “home base” of OSM users (local mappers). This is necessary because this information is not saved within the OSM user data or cannot be read simply from IP addresses [5]. On the other hand, methods seek to find out where the density of local information is high (local content).

Since other scholars mostly concentrate on either of the two concepts, we are most interested in contrasting both approaches. We do so by implementing two methods that work on a global scale – one focusing on “local” mappers and one aiming at identifying “local” content. Using the rather simple method of looking at the first OSM changeset of a user and similar to Neis [6], we examine the “localness” of mappers across the world. Since it is always possible that a new OSM user first has mapped remotely or away from home on a trip, this method can therefore only serve as an approximation of fundamental global inequalities. An advantage of exploring the first changeset is that it can easily be applied globally to all and not exclusively to very active mappers.

To focus on the objects in OSM, we proceed with a method that Zielstra et al. [7] have already applied at a regional level for selected mappers. In order to identify areas with a higher density of local content, we use the framework from Raifer et al. [8] to calculate

Schröder-Bergen, S. (2020). Analyzing the localness of OSM data
In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2020 Online Conference, July 4-5 2020. Available at https://zenodo.org/communities/sotm-2020
DOI: 10.5281/zenodo.3923062

multiple indicators on a global level. We determine the ratio of the sum of all used OSM tags to the number of OSM elements per spatial unit. Furthermore, the ratio of the sum of the different OSM keys to the number of OSM elements per spatial unit is computed. It is assumed that a larger number of tags and a greater variability within the keys is most likely to be associated with a higher “localness”.

On a global level, these two approaches thus attempt a data-driven operationalization and approximation of “localness”. Of particular interest is the extent to which local content actually coincides with the distribution of local mappers. Based on the state of research on social inequalities in VGI and OSM, it can be expected that “localness” follows a regional center-periphery and a global north-south divide. However, it is also to be assumed that other phenomena besides “localness” will be observed. For this, a look beyond the users to the OSM objects and their variability is heuristically valuable. An analysis of the tags provides further information about the richness of the OSM data, which indicates the involvement of local knowledge.

Our research could be followed up by a case study on “localness” in a specific regional context. We suggest that both active user groups have to be considered and reference should be made to the OSM object level. In doing so, it is reasonable to dive into exemplary communities in which the contested meaning of local knowledge becomes evident. At the object level, the role of “localness” in relation to the richness and variability of the data could be examined. However, the research conducted so far has provided a more fundamental answer to the question in what way locally produced data in OSM are different to remote data. Being considered as an important indicator for the authenticity of data in OSM, this research helps to gain a better understanding of “localness” in OSM.

References

[1] Goodchild, M. F. (2007). Citizens as sensors. The world of volunteered geography. *GeoJournal,* 69(4), 211-221.

[2] Elwood, S. A., Goodchild, M. F., & Sui, D. Z. (2012). Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the American Association of Geographers*, 102(3), 571-590.

[3] Barron, C., Neis, P., & Zipf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, 18(6), 877-895.

[4] Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2016). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31(1), 139-167.

[5] Quinn, S. (2016). A Geolinguistic Approach for Comprehending Local Influence in OpenStreetMap. *Cartographica*, 51(2), 67-83.

[6] Neis, P. (2013). The OpenStreetMap Contributors Map aka Who’s around me? Retrieved from http://neis-one.org/2013/01/oooc.

[7] Zielstra, D. W. G., Hochmair, H., Neis, P., & Tonini, F. (2014). Areal delineation of home regions from contribution and editing patterns in OpenStreetMap. *IJGI*, 3(4), 1211-1233.

[8] Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., Przybill, K., Fendrich, S., Mocnik, F.-B., & Zipf, A. (2019). OSHDB: a framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospatial Data, Software and Standards*, 4(1).

Evolution of humanitarian mapping within the OpenStreetMap Community

Benjamin Herfort[1,2,*], Martin Raifer[1], Jochen Stier[1], Alina Klerings[1] and Marcel Reinmuth[1]

[1] Heidelberg Institute for Geoinformation Technology, affiliated Institute at Heidelberg University, Heidelberg, Germany; benjamin.herfort@heigit.org, martin.raifer@heigit.org, j.stier@stud.uni-heidelberg.de, klerings@stud.uni-heidelberg.de, marcel.reinmuth@stud.uni-heidelberg.de
[2] GIScience Research Group, Heidelberg University, Heidelberg, Germany;

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2020 Online Conference after peer-review.

Since 2010 organized humanitarian mapping has evolved as a constant and growing element of the global OpenStreetMap (OSM) community. In the last few years, several researchers have analyzed humanitarian mapping practices [1, 2, 3], however most of this work was either event-driven or focused on specific time periods and regions. The OSM ecosystem and the actors involved in it are constantly changing and emerging, for instance also due to the mapping activities of corporates [4]. In our work we analyze the history of almost 10 years of humanitarian mapping in OpenStreetMap using the OSM Tasking Manager (tasks.hotosm.org). We conduct a comprehensive quantitative analysis on a global scale and long term perspective in order to depict more than just snapshots of individual events. We show how humanitarian mapping was impacted by major mapping disaster response events, but also widened its application to other domains such as disaster preparedness. Our approach follows two paths. One focuses on the mapping itself. The other focuses on the composition of the humanitarian mapping community.

Our analytical approach is characterized by the combination of the following data sets: for general OSM mapping and user characteristics we use the whole history of OSM object versions and edits provided by the OSHDB framework [5]. We were provided with a HOT Tasking Manager database dump covering the time between 11/2012 until 12/2019. Our method uses spatio-temporal time series analysis based on a hexagonal grid for all areas humanitarian mapping activity has been conducted. We investigate the evolution of humanitarian mapping with respect to several attributes including number of OSM contributions (created, modified, deleted OSM objects), percent of tasks mapped and validated in the Tasking Manager and number of OSM contributors involved. We set our findings and the geographical distribution of our results in context to the overall OSM mapping and user characteristics.

Humanitarian mapping has grown almost linear with respect to the numbers of projects, unique users, tasks mapped and OpenStreetMap contributions, while at the same time it got more diverse in terms of spatial distribution of mapping activities and organizations involved. In 2018 more than 100 organizations were running projects on every

Herforth, B., Raifer, M., Stier, J., Klerings, A., & Reinmuth, M. (2020). Evolution of humanitarian mapping within the OpenStreetMap Community
In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2020 Online Conference, July 4-5 2020. Available at https://zenodo.org/communities/sotm-2020
DOI: 10.5281/zenodo.3923064

inhabited continent except Australia. Peaks in mapping activities in the first quarter of 2015 and the third quarter of 2017 are reflecting activations of the community in responses to the Gorkha earthquake as well as to the hurricanes Harvey, Irma and Puebla earthquake. Although the number of mappers is growing, the humanitarian community lacks long term commitment. Almost 70% of humanitarian mappers contribute one day only. Another 10% drops out after ten days of being active. However the number of one day only contributors is steadily rising, whereas mappers involved on 5 or more days seem to be saturated since 2017. Nonetheless, the one day only contributors have a considerable impact on the data. They account for 10% of all edits, whereas contributors involved on 100 or more days account for 50%. This pattern is reflected by the composition of user actions as well. The number of mapping activities is growing almost linear, whereas the amount of validators stagnate. However, different organisations have varying degrees of success in recruiting new validators. The American Red Cross for instance accounts for 30 to 50% of all first time validators in the last 12 months.

Humanitarian mapping via the tasking manager now exists for more than seven years. In terms of scale it is definitely a success story. We have shown edits, users, projects, geographic diversity, almost all of these have experienced linear growth. But with respect to the numbers of user commitment and validation efforts we conclude that the humanitarian mapping community still faces huge challenges to achieve sustainability. Only a small proportion of users contribute regularly and an even smaller fraction does validation. But this very phenomenon is nothing new for OSM in general. Nevertheless, we have seen that some organisations and communities are more successful than others in recruiting and retaining users. Our results may support decisions for future strategies on user engagement. However to what degree this community composition affects data quality remains open. Within our work our definition of humanitarian mapping is rather narrowed. Our insights about humanitarian mapping in OSM provide only an incomplete picture which lacks an on-the-ground perspective and neglects other remote mapping tools.

References

[1] Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. *World Medical & Health Policy*, 2(2), 7-33.

[2] Soden, R., & Palen, L. (2016). Infrastructure in the Wild: What Mapping in Post-Earthquake Nepal Reveals about Infrastructural Emergence. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose, California, USA, 2796-2807.

[3] Dittus, M., Quattrone, G., & Capra, L. (2017). Mass Participation During Emergency Response: Event-centric Crowdsourcing in Humanitarian Mapping. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, Portland, Oregon, USA, 1290-1303.

[4] Anderson, J., Sarkar, D., & Palen, L. (2019). Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 8(5), 232.

[5] Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., Przybill, K., Fendrich, S., Mocnik, F.-B., & Zipf, A. (2019). OSHDB: A framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospatial Data, Software and Standards*, 4(3).

Detecting informal settlements via topological analysis

Satej Soman[1], Cooper Nederhood[1], Nicholas Marchio[1], Annie Yang[1], Anni Beukes[1,*] and Luis Bettencourt[1,2,3]

[1] Mansueto Institute for Urban Innovation, University of Chicago, Chicago, IL, USA;
satej@uchicago.edu, cnederhood@uchicago.edu, nmarchio@uchicago.edu, ayang20@uchicago.edu, beukes@uchicago.edu, bettencourt@uchicago.edu
[2] Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA;
[3] Sante Fe Institute, Sante Fe, NM, USA;

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2020 Online Conference after peer-review.

Determinations of whether a neighbourhood is underserviced depends greatly on local concerns, however topological analysis of building footprints in relation to streetlevel access offers a pathway to near-universal criteria for determining whether a neighbourhood is potentially a slum or urban informal settlement. With a topological approach, it is possible to determine the number of building parcels a streetblock inhabitant must cross to access a potential services bearing street, i.e. one that enables for example access to emergency services, sanitation and piped water. Focusing on topological invariants allows us to analyse cities without respect to the specific morphology of their street network [1]. To create a global index of these under-serviced neighbourhoods, we extracted open data on building footprints and street networks and applied a topological analysis to each extracted street block to characterise the level of spatial accessibility. Using these techniques, we analysed 1 terabyte of OpenStreetMap (OSM) data to create an index of street block geometry, cadastral maps, and spatial accessibility calculations for 120 Low and Middle Income Countries (LIMCs) in the Global South. These results highlight the developing world's most spatially inaccessible communities, enabling prioritisation of infrastructure investment and further study.

The primary data universe for our work involves three entities: 1) administrative boundary polygons, 2) building footprint polygons, and 3) street network line-string collections. The administrative data boundaries were sourced from the Database of Global Administrative Areas, while building footprints and street networks were sourced from OSM. The OSM data used was obtained from a mirror of OSM hosted by GeoFabrik. This data reflects a snapshot of OSM as of July 2019.

Starting with the initial road network, the geometry of the street network can be extracted via the well-studied process of polygonization. In polygonization, the self-intersections of the street network are determined in order to define the block geometry. Most GIS software packages offer a polygonization feature (e.g. *ST_MakePolygon* from

Soman, S., Nederhood, C., Marchio, N., Yang, A., Beukes, A., & Bettencourt, L. (2020). Detecting informal settlements via topological analysis
In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2020 Online Conference, July 4-5 2020. Available at https://zenodo.org/communities/sotm-2020
DOI: 10.5281/zenodo.3923066

PostGIS). In implementation, however, we found a set-theoretic approach to be more stable and performant [2].

By buffering the one-dimensional line-strings comprising the street network with a small buffer radius, we obtain two-dimensional polygons capturing the outline of the street network. We then find the set-theoretic difference between the administrative boundary polygons and the buffered road-network polygon; this renders the negative area between the road network geometry as a collection of polygons. These negative areas are precisely the street blocks we are trying to extract.

With each building's access to the formal road network – or lack thereof – analysed, we can also provide suggestions to the local community about ways to connect each building to the existing road network while respecting existing building footprints. This process, known as re-blocking, is in general non-deterministic polynomial-time hard (NP-hard) and there is therefore not an efficient algorithm offering a solution to the problem. In contrast to previous algorithmic re-blocking efforts involving stochastic graph search [3], we approach the problem using Steiner tree approximations to the optimal road network.

To frame the universal access network as a Steiner tree problem, we segment the building parcel boundaries to create "non-terminal" nodes at each segment boundary, and create a "terminal" node for each building parcel lacking direct street access by placing a node on the non-street parcel boundary closest to the building. We then construct a complete subgraph of all the terminal nodes, where the weight of each edge is the Euclidean distance between each pair of terminal nodes; each existing road segment has zero weight. Solving the minimum-spanning tree problem on this subgraph, and then recovering the original path segments, gives us the optimal new road network.

In the short future, we hope to repeat our analysis on denser sources of data on footprints and street network data from other providers (such as private satellite imagery companies). We expect that while OSM has good coverage in some areas, we can use data in which footprints are automatically extracted from satellite imagery to supplement areas where coverage may be lacking. Additionally, we hope to re-block more neighbourhoods and cities and investigate the connections between an urban centre's infrastructural topology and the amount of new infrastructure needed to be built to universally connect each building to the formal road network. Finally, we expect that our generated dataset could be used to study development and health outcomes, especially around questions of tenure security and emergency service provision.

References

[1] Brelsford, C., Martin, T., Hand, J., & Bettencourt, L. M. A. (2015). The Topology of Cities. *SFI Working Paper 2015-06-021*. Retrieved from https://tinyurl.com/ybcdpktz.

[2] Geographic Information Systems Stack Exchange (2013). Generate Polygons from a Set of Intersecting Lines. Retrieved from https://tinyurl.com/y9va53lh.

[3] Brelsford, C., Martin, T., & Bettencourt, L. M. A. (2019). Optimal reblocking as a practical tool for neighborhood development. *Environment and Planning B: Urban Analytics and City Science*, 46(2), 303-321.