

Analyzing the spatio-temporal patterns and impacts of large-scale data production events in OpenStreetMap

A. Yair Grinberger^{1,2,*}, Moritz Schott², Martin Raifer², Rafael Troilo² and Alexander Zipf²

¹ Department of Geography, The Hebrew University of Jerusalem, Israel; yair.grinberger@mail.huji.ac.il

² GIScience Research Group, Heidelberg University, Heidelberg, Germany;

yair.grinberger@mail.huji.ac.il, M.Schott@stud.uni-heidelberg.de, martin.raifer@uni-heidelberg.de, rafael.troilo@uni-heidelberg.de, zipf@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.

Volunteered geographical information often visions data as a product of individual actions [1]. In OpenStreetMap (OSM) however, contributions are frequently made as part of large-scale data production events [2, 3]. These events, which can take multiple forms, do contribute much to the OSM project. Nevertheless, they also hold the potential to significantly affect the map by changing the development course of data and community in accordance with the perspectives of the event's organizers and participants. Hence, it is important to identify and understand such events, as well as their impacts upon the data.

This study sets out to contribute to the study of these issues by analyzing their spatio-temporal patterns and impacts, based on a novel procedure for automatically identifying large-scale events and classifying them. The identification procedure relies on the assumption, derived from the model proposed by [4], that in the absence of interventions the cumulative distribution of contribution actions (i.e. the operations made as part of each contribution, be it a creation, deletion, or edit) would be S-shaped. This, since data grows exponentially as the community grows until it reaches some form of saturation. Accordingly, we fit an S-shaped logistic curve to the cumulative distribution of contribution actions over time for different regions, computed per month out of the full history of OSM using the OSM History Database tool (OSHDB) [5]. We identify events where the increase in the cumulative number of actions is significantly higher than predicted. Thus, events are defined in terms of both their absolute size and their relative weight in the development of the data. Events are classified via a clustering procedure relying on measures which represent the centralization of events and their contribution themes, i.e. the maximal share of contributions made by one user and the share of different contribution types out of all contributions.

The results show that a significant share of all OSM contributions are made as part of an event, with some data regions almost entirely dominated by these. Furthermore, it does not seem that over the years the role of events in producing a significant share of all new

Grinberger et al. (2019). Analyzing the spatio-temporal patterns and impacts of large-scale data production events in OpenStreetMap

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, 9-10. Heidelberg, Germany, September 21-23, 2019. Available at <https://zenodo.org/communities/sotm-2019>

DOI: [10.5281/zenodo.3387671](https://doi.org/10.5281/zenodo.3387671)



contribution actions has significantly diminished. Looking deeper into the nature of events, we identify two different event types based on the contribution of individuals – local events and remote mapping events – and several bulk import event types, diverging mostly in the share of creations in the events' contributions. Computing the number of events over time shows that while data creation imports were the most frequent type of events early on, over the last years remote mapping events are contributing the most data. Locally based events also show a significant increase in data production. However, these types of events are not distributed evenly across the globe, with import events frequent mostly in countries with developed economies and remote mapping events being more common in the least developed regions of the world. Interestingly, and in contrast with logical expectations, for remote mapping events there is no clear correlation between the timing of an event (i.e. how early it took place) and the share of its contributions out of the current total number of contributions. This expectation is true however for import events. Hence, mapping and analyzing large-scale events allows relating the nature of representation to socio-economic effects.

This study would further break down the spatio-temporal patterns of events, investigating whether the temporal patterns for different regions follow the global ones or are there clusters of temporal change as well. Furthermore, the talk would study the nature of events' impacts, presenting how the values of measures such as the stability of events' contributions and change in the number of active mappers vary by event type and area. These results, beyond promoting a deeper understanding of events and representation in OSM, would allow assessing the implications for the project of current and expected trends in OSM data production. This would allow identifying types of events that were successful, e.g. had led to the enrichment of the data and/or contributed to the size and diversity of the community, hence helping in formulating general guidelines for large-scale events which also consider the local context such as the state of the data and community.

References

- [1] Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211-221.
- [2] Palen, L., Soden, R., Anderson, T. J., & Barrenechea, M. (2015). Success & scale in a data-producing organization: The socio-technical evolution of OpenStreetMap in response to humanitarian events. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 4113-4122.
- [3] Zielstra, D., Hochmair, H. H., & Neis, P. (2013). Assessing the Effect of Data Imports on the Completeness of OpenStreetMap—A United States Case Study. *Transactions in GIS*, 17(3), 315-334.
- [4] Gröchenig, S., Brunauer, R., & Rehl, K. (2014). Digging into the history of VGI data-sets: results from a worldwide study on OpenStreetMap mapping activity. *Journal of Location Based Services*, 8(3), 198-210.
- [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), 1-12.