Towards understanding the temporal accuracy of OpenStreetMap: A quantitative experiment

Levente Juhász[1,*]

[1] GIS Center, Florida International University, Miami, FL, USA; ljuhasz@fiu.edu

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

The ability to provide timely information compared to traditional collection methods of geographic information is generally considered as one of the main advantages of volunteered geographic information (VGI) since its emergence in the 2000s [1]. In addition to several anecdotal examples illustrating how VGI data can provide more up-to-date information than authoritative sources, the literature provides ample evidence on the usefulness of VGI in applications that require timely geodata, such as disaster management [2, 3]. For example, the Haiti earthquake relief effort in 2010 laid the foundations for how remote contributors of OpenStreetMap (OSM) and other platforms can make a difference and aid responding humanitarian agencies after a crisis [4]. The Humanitarian OpenStreetMap Team has made numerous contributions and helped save lives at numerous instances ever since [5]. However, apart from these examples, the temporal dimension of VGI has not received much research attention outside the application of disaster management, and there is a huge gap between assessing temporal accuracy and other factors of data quality, such as spatial accuracy [6, 7]. Abreicht et al. highlighted the lack of formal acknowledgment of temporal aspects in the concept of VGI and proposed a framework called 'Volunteered Geo-Dynamic Information' to fully integrate spatial and temporal aspects of VGI [8]. Other works utilizing the temporal component in VGI often focus on the behavior of contributors rather than the currency and temporal validity of map features they contributed [9–11], or studied the evolution of data over time [12, 13]. While these approaches are useful, by nature they cannot provide a quantitative measure of how current OSM (or VGI in general) is. [14] noted during their investigations that the temporal accuracy of OSM could not be measured using their traditional extrinsic method, because OSM data was compared to authoritative data that did not contain temporal information (i.e. most recent street configuration regardless of when road segments were built or renovated). Another project, 'Is OSM up-to-date?' recognizes the lack of information on temporal accuracy and developed a tool that uses an intrinsic approach to visually show features that potentially contain outdated information [15]. However, by nature, an intrinsic approach can also not provide an absolute measure of how up-to-date OSM is.

This research attempts to fill a gap in the literature by conducting an experiment on the currency of VGI. Using OSM data as a case study, it measures the temporal accuracy of

Juhász, L. (2021). Towards understanding the temporal accuracy of OpenStreetMap: A quantitative experiment
In: Minghini, M., Ludwig, 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, 19-22. Available at https://zenodo.org/communities/sotm-2021
DOI: 10.5281/zenodo.5112236

selected map features. This research overcomes previous limitations by using official data provided by the Florida Department of Transportation (FDOT). The dataset contains details about state-funded highway construction projects, including the date these projects were completed, therefore, accurately measuring the temporal accuracy of OSM features is possible by comparing dates projects were finished with the time at which corresponding OSM edits in the database were made. This time difference describes how long it took the OSM community to adapt to real-world changes and update the map database accordingly.

The historical version of highway construction projects was filtered to projects completed between May 15, 2016 and April 1, 2021. Further, only a subset of projects were used, that resulted in either new infrastructure (new roadways, roundabouts or highway ramps) or widening of existing roadways (i.e. new lanes excluding bike lanes and turning lanes). Other construction projects, such as traffic improvements, road resurfacing, regular maintenance (e.g. bridge rehabilitation), bike infrastructure etc. were excluded, since a useful, high-quality road network database can be maintained without the addition of this information, therefore, they are less likely to migrate into OSM. The methodology uses augmented diffs from the Overpass API to find all changes that occurred on OSM highway features (creation, modification and deletion) and are spatially and temporally close to construction projects. These changes are then matched with a record from the highway construction dataset. Irrelevant changes (i.e. changes made to other highway features) are removed. This is done by manually interpreting and evaluating changes and construction projects using a description field (e.g. *"WAKULLA SPRINGS ROAD @ OAK RIDGE ROAD ROUNDABOUT"*). The data extraction algorithm initially queries the Overpass API for changes one week beyond the completion date of a particular project. In case no relevant change can be found, iterative queries for 7-day-long time slices are made until a relevant change is found, or until the current date is reached. Lastly, the time difference between the end date of construction projects and the first OSM change that introduced the change in OSM are calculated. For example, the example from above can be found with the following Overpass query (https://overpass-turbo.eu/s/16XV) that uses the location of the highway construction. The relevance of a change can be verified using changeset comments: (https://www.openstreetmap.org/changeset/87938707). In this example, the changeset comment *"Added new round about."* confirms that the OSM edit is related to the FDOT dataset. The difference between the construction end date (July 3, 2019) and the time when this change appeared in OSM (July 13, 2020) is 1 year and 7 days.

The FDOT construction dataset was manually checked and matched with OSM edits. The final dataset contains 23 new highways and roundabouts, and 44 road widening projects (lane additions). Only 3 out of 23 highway construction projects have not been added to OSM. However, these projects were recently finished (October 2020), therefore it is possible that they will be added to OSM at a later time. The remaining projects were mapped in OSM within 133 days on average (median=45, SD=262). Widening projects are less likely to be mapped in OSM in a timely fashion, as 23 out of 44 of those projects are not mapped. The remaining widening projects seem to be added at a slower rate (mean=158 days, median=147, SD=388). Boxplots of construction projects are shown in Figure 1. Interestingly, several projects of both types have been mapped in OSM ahead of the official construction end date. This shows the flexibility of the OSM, which is able to adapt to informal scenarios, where roadways are already being used while construction crews are still working.

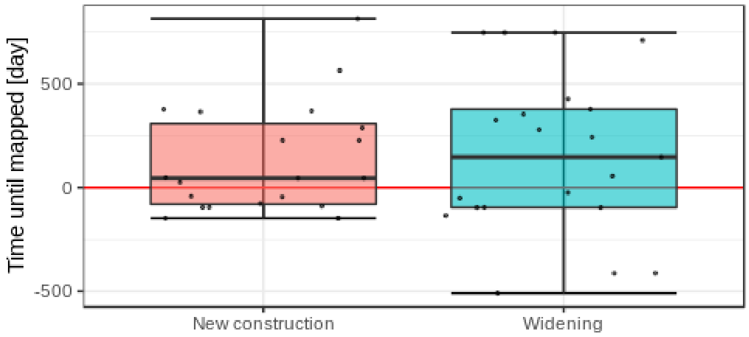

Figure 1. Lag between construction projects and corresponding mapping events

This experiment is the first attempt to investigate the timeliness and currency of VGI using OSM as a case study. The limitations of the study include the reference dataset, that does not contain federally or locally funded projects, therefore misses a large number of constructions, and the methodology, that cannot capture the diversity of the OSM community and also disregards changes beyond the transportation infrastructure. Future work will conduct analysis using more VGI data sources outside the domain of mapping applications (e.g. Points of Interest), new methodology using tile-reduce, OSM QA tiles and vector tiles built from other datasets. The new methodology will be scalable and will allow for analysis across world regions. Furthermore, a rule-based decisions approach based on tags and semantics will be used to eliminate the need for manually checking and verifying whether VGI updates correspond to the reference dataset or not.

References

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

[2] Horita, F. E., Degrossi, L., Assis, L. F. de, Zipf, A., & de Albuquerque, J. P. (2013). The use of Volunteered Geographic Information (VGI) and Crowdsourcing in Disaster Management: A Systematic Literature Review. In: *Proceedings of the 19th Americas Conference on Information Systems*.

[3] Neis, P., & Zielstra, D. (2014). Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap. *Future Internet*, 6(1), 76-106.

[4] 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.

[5] Herfort, B., Lautenbach, S., de Albuquerque, J. P., Anderson, J., & Zipf, A. (2021). The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific Reports*, 11(1), 3037.

[6] Antoniou, V., & Skopeliti, A. (2017). The Impact of the Contribution Microenvironment on Data Quality: The Case of OSM. In: Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C. C., & Antoniou, V. (Eds.) *Mapping and the Citizen Sensor*, 165-196. Ubiquity Press, London.

[7] 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*, 34(9), 1765-1791.

[8] Aubrecht, C., Aubrecht, D. Ö., Ungar, J., Freire, S., & Steinnocher, K. (2017). VGDI – Advancing the Concept: Volunteered Geo-Dynamic Information and its Benefits for Population Dynamics Modeling. *Transactions in GIS*, 21(2), 253-276.

[9] Begin, D., Devillers, R., & Roche, S. (2018). The life cycle of contributors in collaborative online communities -the case of OpenStreetMap. *International Journal of Geographical Information Science*, 32(8), 1611-1630.

[10] Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal*, 47(4), 315-322.
[11] 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.
[12] Girres, J.-F., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4), 435-459.
[13] Zielstra, D., & Hochmair, H. H. (2011). A Comparative Study of Pedestrian Accessibility to Transit Stations Using Free and Proprietary Network Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2217, 145-152.
[14] Arsanjani, J. J., Barron, C., Mohammed, B., & Helbich, M. (2013). Assessing the Quality of OpenStreetMap Contributors together with their Contributions. In: Vandenbroucke, D., Bucher, B., & Crompvoets, J. (Eds.) *Proceedings of the 16th AGILE Conference on Geographic Information Science*, 14-17.
[15] Minghini, M., & Frassinelli, F. (2019). OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date? *Open Geospatial Data, Software and Standards*, 4(1), 9.