Detection and Semantic Annotation of Changes in OpenStreetMap Data

Camille Bernard[1,*], Pedro Pinheiro[1] and Matthieu Viry[2]

[1] University Grenoble Alpes, CNRS, Grenoble INP, LIG, Grenoble, France; camille.bernard@univ-grenoble-alpes.fr, pedro-henrique.santos-pinheiro@grenoble-inp.org
[2] Université Paris Cité, CNRS, Centre pour l'analyse spatiale et la géovisualisation, F-75013 Paris, France; matthieu.viry@cnrs.fr

Geographical databases such as OpenStreetMap (OSM) continuously evolve to reflect ground-level transformations. Characterizing these changes is key to analyzing territorial evolution over time. Previous studies have addressed this evolution from different perspectives, including the evolution of OSM data quality and completeness: [1] established one of the first frameworks for quality assessment based on OSM's version history, [2] and [3] examined the growth and structural evolution of OSM road networks as a proxy for completeness, and [4] proposed a broader framework combining temporal indicators to capture evolving data reliability. Other works have studied contributor behavior and editing dynamics [5, 6, 7], including the detection of anomalous or malicious edits such as vandalism [8, 9]. On the semantic side, [10] showed how OSM tags emerge, consolidate, and stabilize over time, while infrastructures such as the OpenStreetMap History Database [11] provide tools for OLAP-style analysis of OSM's editing history.

Despite this body of work, the semantic interpretation of OSM changes (in particular, the ability to distinguish genuine ground-level evolution from data refinements, tagging corrections, or modeling decisions) remains barely unaddressed. [12] offers a first step in this direction, introducing a provenance model for OSM that analyzes collaborative editing patterns to describe how features evolve through successive contributions. We build on this foundation and extend it further: we propose a richer, more generic vocabulary (the GeoChange Ontology https://geochange-ontology-5d3499.gricad-pages.univ-grenoble-alpes.fr/) that captures not only editing operations but the underlying editing intention (e.g., error correction, error introduction, data model refinement, or real-world change). While generically designed to address any type of geographic dataset, including authoritative data, we focus here on OSM as a particularly relevant case, given its collaborative and continuously evolving nature. For planners relying on OSM, distinguishing real infrastructure changes from editing artifacts is critical.

We present a framework for analyzing temporal changes in OpenStreetMap data through a two-level approach. The first level, currently operational, detects and classifies observable feature transitions based on four dimensions: existence, tag change, geometry change, and structural change. The second level, currently under development, aims to interpret these transitions by reasoning over groups of changes using spatio-temporal coherence, in order to distinguish real-world changes from data-related modifications. The approach is applied to cycling infrastructure datasets in three study areas, and is complemented by a web application for exploring editing patterns and visually inspecting feature evolution over time. These results provide a foundation for a scalable framework dedicated to OSM change interpretation, with future integration of semantic annotations into a Knowledge Graph to support urban monitoring and data quality assessment.

This work was funded by the French National Research Agency through the GEVOK project (ANR-25-CE23-6986 2026-2029).

References

[1] Girres, J.-F., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS, 14*(4), 435–459. https://doi.org/10.1111/j.1467-9671.2010.01203.x

[2] Neis, P., Zielstra, D., & Zipf, A. (2012). The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. *Future Internet, 4*(1), 1–21. https://doi.org/10.3390/fi4010001

[3] Corcoran, P., Mooney, P., & Bertolotto, M. (2013). Analysing the growth of OpenStreetMap networks. *Spatial Statistics, 3*, 21–32. https://doi.org/10.1016/j.spasta.2013.01.002

[4] Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS, 18*(6), 877–895. https://doi.org/10.1111/tgis.12073

[5] 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. https://doi.org/10.3390/ijgi1020146

[6] Mooney, P., & Corcoran, P. (2012). The annotation process in OpenStreetMap. *Transactions in GIS, 16*(4), 561–579. https://doi.org/10.1111/j.1467-9671.2012.01306.x

[7] Mooney, P., & Corcoran, P. (2014). Analysis of interaction and co-editing patterns amongst OpenStreetMap contributors. *Transactions in GIS, 18*(5), 633–659. https://doi.org/10.1111/tgis.12051

[8] Neis, P., Goetz, M., & Zipf, A. (2012). Towards automatic vandalism detection in OpenStreetMap. *ISPRS International Journal of Geo-Information, 1*(3), 315–332. https://doi.org/10.3390/ijgi1030315

[9] Quinn, S., & Bull, F. (2019). Understanding Threats to Crowdsourced Geographic Data Quality Through a Study of OpenStreetMap Contributor Bans. In Geospatial Information System Use in Public Organizations. Routledge.

[10] Mocnik, F.-B., Zipf, A., & Raifer, M. (2017). The OpenStreetMap folksonomy and its evolution. *Geo-Spatial Information Science*, *20*(3), 219–230. https://doi.org/10.1080/10095020.2017.1368193

[11] 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), 3. https://doi.org/10.1186/s40965-019-0061-3

[12] Keßler, C., Trame, J., & Kauppinen, T. (2011). Tracking editing processes in volunteered geographic information: The case of OpenStreetMap. In *Identifying objects, processes and events in spatio-temporally distributed data (IOPE), workshop at conference on spatial information theory*.