OpenStreetMap as an emerging scientific field: Reflections from OSM Science 2023

A. Yair Grinberger[1*], Hao Li[2], Pengyuan Liu[3], Godwin Yeboah[4], Levente Juhasz[5], Serena Coetzee[6], Peter Mooney[7], Alessandro Sarretta[8], Jennings Anderson[9], and Marco Minghini[10,†]

[1] Department of Geography, The Hebrew University of Jerusalem, Israel; yair.grinberger@mail.huji.ac.il
[2] Department of Aerospace and Geodesy, Technical University of Munich, Ottobrunn, Germany; hao_bgd.li@tum.de
[3] School of Geographical Sciences, Nanjing University of Information Science and Technology, China; neil.pengyuanliu.0705@gmail.com
[4] Research Computing, Research Technology Platforms, 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 Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa; serena.coetzee@up.ac.za
[7] Department of Computer Science, Maynooth University, Maynooth, Ireland; peter.mooney@mu.ie
[8] Research Institute for Geo-Hydrological Protection, National Research Council, Padova, Italy; alessandro.sarretta@cnr.it
[9] Meta Platforms Inc., Salem, OR, United States; jenningsa@fb.com
[10] European Commission, Joint Research Centre (JRC), Ispra, Italy; marco.minghini@ec.europa.eu

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

OpenStreetMap (OSM) started as a project in 2004 aiming at creating a digital and open map of the world via collaborative mapping, emerging over time to become a community (or a collection of communities) [1] or an ecosystem [2] around the project itself. This ecosystem encompasses local and global communities of data and software developers creating a large number of tools and services, e.g. for spatial data infrastructures [3], disaster response [4], and routing [5]. Additionally, a new scientific field focusing on OSM is emerging with academic researchers investigating the different scientific aspects of the living OSM community [6–8]. The Academic Track at the annual State of the Map (SotM) conference, with five editions from 2018 to 2022 has become a knowledge hub for gathering and sharing recent progress in OSM-related research and scientific findings directly with the broad OSM community. Moreover, the 2019 and 2020 editions of this Track have led to the first special issue of scientific articles dedicated to OSM [9], which first formalized and used the term "OSM Science" outside informal conversations in SotM and the OSM-science mailing list [10] (with one exception being Haklay's reference to 'OSM studies' [11]). In its sixth edition, the Academic Track starts to use the new name of "OSM Science" referring to

Grinberger, A.Y., Li, H., Liu, P., Yeboah, G., Juhász, L., Coetzee, S., Mooney, P., Sarretta, A., Anderson, J., & Minghini, M. (2023). OpenStreetMap as an Emerging Scientific Discipline: Reflections from the OSM Science 2023
In: Minghini, M., Li, H., Grinberger, A.Y., Liu, P., Yeboah, G., Juhász, L., Coetzee, S., Mooney, P., Sarretta, A., & Anderson, J. (Eds.). Proceedings of the OSM Science at State of the Map Europe 2023, Antwerp, Belgium, 10-12 November 2023. Available at https://zenodo.org/communities/osmscience-2023
DOI: 10.5281/zenodo.10443298

the emerging scientific discipline characterized by its unique focus on the OSM project, data, contributors, community, and applications. The proceedings of the OSM Science at the SotM Europe 2023 conference, taking place in Antwerp, Belgium on November 10-12, 2023 [12], include 17 short papers corresponding to 8 talks and 9 lightning talks presented at the conference. In this Editorial, we survey these papers by grouping them into diverse but also interrelated research topics following an interdisciplinary perspective [13].

The study of data quality in OSM has a profound record, since the “fitness-for-use” of OSM data may vary in different application scenarios and spatial-temporal contexts. Several abstracts included in these proceedings concern the topic of OSM data quality assessment. The first two investigate the completeness of OSM buildings and the attribute completeness of OSM roads, respectively. Oostwegel et al. [14] present a dynamic assessment of the completeness of OSM buildings on a global scale, which can be updated at the moment that a building is added, modified, or removed. Similarly, but for a city scale, Andorful et al. [15] explore the association of OSM road classes and Points-of-Interest (POI) categories in their proximity. The key findings are envisioned to guide mappers by suggesting road classes in scenarios where POIs are mapped before the roads themselves to ensure a simultaneous quality check of both road classes and POIs. Beyond buildings and roads in OSM, O’Brien [16] seeks to enhance the quality and completeness of OSM land use information with open datasets, such as governmental retail areas and geolocated energy performance data, in Great Britain. Inspired by the Nolli map, Scheck et al. [17] design a framework to map urban public spaces by sequentially analyzing both relevant tags and geometries of the OSM, then validate and enrich the data via extensive on-site mapping in the city of Vienna, Austria. All these four works emphasize the importance of deriving an up-to-date and accurate assessment of OSM data quality in order to better understand how OSM data can be used.

In addition to the existing literature on data-centric research in the OSM scientific community, a large number of abstracts (7 abstracts) focus on domain-specific applications of OSM data. These include harnessing OSM data to build a baseline for healthcare facilities data [18], 3D city modeling and building height estimation by combining OSM data with remote sensing and street-view imagery [19], and exploring OSM data to measure the level of local proximity to services by walking [20]. Other studies explore the potential for integrating OSM data with other data sources in diverse application scenarios. For instance, Evaz Zadeh et al. [21] highlight OSM potential in improving the accuracy of earthquake risk assessment by estimating building type information from existing OSM building footprints. Vierø et al. [22] relate to the use of OSM for planning bicycle networks, developing the BikeDNA tool which integrates extrinsic and intrinsic data quality analysis procedures to assess bicycle path mapping in OSM (and thus also relate to the theme of data quality). Finally, Passmore et al. [23] assess the accessibility of bike-transit (i.e. the use of bicycle as the first and/or the last-mile mode) using an OSM road network, highlighting the need to improve OSM road tags to enhance bicycle routing algorithms for cycling, walking, and transit use.

More recently, Machine Learning (ML) as an emerging topic has gained quite some momentum in the OSM research community, especially in conjunction with Geospatial Artificial Intelligence (GeoAI). Five abstracts in these proceedings contribute to different aspects of GeoAI with OSM. First, Fila et al. [24] compare the editing behavior pattern of AI-assisted and traditional mapping methods by mining historical OSM user contribution data. Vestena et al. [25] propose a framework that combines image semantic segmentation and monocular depth estimation methods on street view images to conduct micromapping for small-scale features, enhancing the completeness and readability of the OSM map. Gramacki et al. [26] showcase a Python library providing procedures to train embedding models based on OSM and additional geospatial data, thus pushing forward discussions regarding standardizing GeoAI research through a set of commonly accepted models and evaluation data. Melanda et al. [27] analyze the (im)balance in OSM highway tags of nine countries, motivated by the need to provide balanced OSM training data for ML tools that do not lead to biased results. Last but not least, Khellaf et al. [28] develop an interesting natural language interface based on ChatGPT for efficiently querying OSM data and finding meaningful combinations of objects and public space, which helps (investigative) journalists for accurate geo-location in the context of information verification.

These three topics — data quality, applications, and ML — represent the data-oriented approach within OSM Science, which exists alongside the social perspective exploring the contexts from which OSM data emerge and the project's societal implications [11]. This latter aspect is assessed in two abstracts. First, Geddes [29] considers how familiarization with OSM, emerging through the inclusion of OSM within water point mapping in Malawi, may contribute to this effort generally. Schröder-Bergen [30] inspects recent changes in the project from a political economy perspective, using a mixed method to understand the implications of OSM emergence as a prominent data source and service and the increasing involvement of paid and organized mapping teams.

One claim raised in the last work is that although OSM has reached a significant level of maturity, or even because of this, it still goes through significant changes that may alter the very foundations of the project. Similarly, the growing interdisciplinary research interest of and about OSM is a positive sign showing that the scientific endeavor termed as OSM Science [9] is further developing and maturing. But, as its object of interest, it still has room to grow. Changes within the project are certainly a source of new scientific developments, as seen in the increasing interest in corporate editors [31] or unprecedented volumes of political vandalism [32]. Yet, there are also developments outside OSM that will surely drive OSM Science forward, e.g. integrating generative AI and large language models such as ChatGPT into geospatial procedures [33], the Overture Maps initiative [34] whose relations with and impacts on OSM are only beginning to unfold, and the increasing academic research on digital transformation and geographical information science (GIScience). Such developments mark OSM as much more than a crowdsourced database but rather a unique phenomenon with no equivalent in the geospatial domain. As such, existing tags such as GIScience and Volunteered Geographic Information may not be enough to capture the research endeavors that will take place as part of the "OSM Science". Such endeavors will certainly extend beyond disciplinary borders and may give rise to new conceptual and theoretical frameworks. It is thus with great excitement that we look to the future of research with and about OSM, hoping that this OSM Science 2023 meeting will be the first of many meetings that will help shape this emerging field and its agenda.

References

[1] Grinberger, A. Y., Minghini, M., Yeboah, G., Juhász, L., & Mooney, P. (2022). Bridges and barriers: An exploration of engagements of the research community with the OpenStreetMap community. *ISPRS International Journal of Geo-Information, 11*(1), 54.

[2] Schott, M. (2020). *The future of working with OSM data*. https://giscienceblog.uni-heidelberg.de/2020/09/10/the-future-of-working-with-osm-data

[3] Minghini, M., Kotsev, A., & Lutz, M. (2019). Comparing INSPIRE and OpenStreetMap data: how to make the most out of the two worlds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-4/W14*, 167–174.
[4] Humanitarian OpenStreetMap Team (2023) *Home*. https://www.hotosm.org
[5] Neis, P., & Zipf, A. (2008). *Openrouteservice. org is three times "open": Combining OpenSource, OpenLS and OpenStreetMaps.* GIS Research UK (GISRUK 08).
[6] Arsanjani, J. J., Zipf, A., Mooney, P., & Helbich, M. (2015). An introduction to OpenStreetMap in geographic information science: Experiences, research and applications. In J. J. Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.) *OpenStreetMap in GIScience* (pp. 1-15). Springer.
[7] Juhász, L., & Hochmair, H. H. (2018). OSM data import as an outreach tool to trigger community growth? A case study in Miami. *ISPRS International Journal of Geo-Information*, 7, 113.
[8] 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 Geographic Information Science*, *34*(9), 1765–1791.
[9] Grinberger, A. Y., Minghini, M., Juhász, L., Yeboah, G., & Mooney, P. (2022). OSM Science—The Academic Study of the OpenStreetMap Project, Data, Contributors, Community, and Applications. *ISPRS International Journal of Geo-Information*, *11*(4), 230.
[10] *Science Info Page*. https://lists.openstreetmap.org/listinfo/science
[11] Haklay, M. (2015). Foreword: OpenStreetMap studies and volunteered geographic information. In J J. Arsanjani, , A. Zipf, P. Mooney, & M. Helbich (Eds.) *OpenStreetMap in GIScience: Experiences, Research, and Applications* (pp. v-vii). Springer.
[12] OSM Science Scientific Committee (2023). *OSMScience 2023*. https://shorturl.at/beAP9
[13] Grinberger, A. Y., Anderson, J., Mooney, P., Ludwig, C., & Minghini, M. (2021). OpenStreetMap as a multifaceted research subject: the Academic Track at State of the Map 2021. In M. Minghini, C. Ludwing, J. Anderson, P. Mooney, & A. Y. Grinberger. (Eds.) *Proceedings of the Academic Track at the State of the Map 2021* (pp. 1–5). Zenodo.
[14] Oostwegel, L. J. N., Evaz Zadeh, T., Lingner, L. & Schorlemmer, D. (2023). A global and dynamic completeness assessment of the OpenStreetMap buildings. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 6–9). Zenodo.
[15] Andorful, F., Lautenbach, S., Ludwig, C., Herfort, B., Nir, F., & Zipf, A. (2023). Exploring Road and Points of Interest (POIs) Associations in OpenStreetMap, A New Paradigm for OSM Road class Prediction. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 69–72). Zenodo.
[16] O'Brien, O. (2023). Towards an Open High-Resolution Land Use Dataset in Great Britain – Comparing and Consolidating Retail Centre Areas from Open Data Sources. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 29–32). Zenodo.
[17] Scheck, E., Ledermann, F., Binn, A. & Dörk, M. (2023). Mapping Public Space in Urban Neighbourhoods Using OpenStreetMap Data. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 33–37). Zenodo.
[18] Herringer, M., Ndiaye, L., & South, A. (2023). Developing a data validation method with OpenStreetMap Senegal and the Ministry of Health in support of accurate health facility data. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 42–44). Zenodo.
[19] Li, H. & Sun, Y. (2023). Beyond Two Dimensions: Large-Scale Building Height Mapping in OpenStreetMap via Synthetic Aperture Radar and Street-View Imagery. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 38–41). Zenodo.

[20] Olivari, B., & Cimini, A. (2023). Are Italian cities already 15-minute? Presenting a glocal proximity index, based on open data. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 57–60). Zenodo.

[21] Evaz Zadeh, T., Oostwegel, L. J. N., Lingner, L., Shinde, S., Cotton, F., & Schorlemmer, D. (2023). Improving the accuracy of earthquake risk estimates with OpenStreetMap building data. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 18–21). Zenodo.

[22] Vierø, A. R., Vybornova, A. & Szell, M. (2023). Assessing OpenStreetMap bicycle data quality with BikeDNA: a Denmark Case Study. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 22–25). Zenodo.

[23] Passmore, R., Guensler, R., & Watkins, K. (2023). Assessing Bike-Transit Accessibility with OpenStreetMap. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 53–56). Zenodo.

[24] Fila, M., Štampach, R., &Benjamin Herfort, B. (2023). Global and regional level of use of buildings and roads prepared by AI for OSM mapping. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 10–13). Zenodo.

[25] Vestena, K., Camboim, S., & Santos, D. (2023). Fostering OSM's Micromapping Through Combined Use of Artificial Intelligence and Street-View Imagery. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 14–17). Zenodo.

[26] Gramacki, P., Leśniara, K., Raczycki, K., Woźniak, S., & Szymański, P. (2023). Utilizing OSM data in geospatial representation learning. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 45–48). Zenodo.

[27] Melanda, E. A., Herfort, B., Ulrich, V., Andorful, F., Zipf, A. (2023). OpenStreetMap Data for Automated Labelling Machine Learning Examples: The Challenge of Road Type Imbalance. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 65–68). Zenodo.

[28] Khellaf, L., Schlicht, I. B., Bayer, J., Bouwmeester, R., Miraß, T., & Tilman Wagner, T. (2023). Spot: A Natural Language Interface for Geospatial Searches in OSM. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 49–52). Zenodo.

[29] Geddes, A. (2023). Rural water point mapping with/in OSM: implications of recent research in Malawi. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 61–64). Zenodo.

[30] Schröder-Bergen, S. (2023). Social, technical and political transformations in OpenStreetMap – From volunteered geographic information to embedding digital commons in platform capitalism. In M. Minghini, H. Li, A. Y. Grinberger, P. Liu, G. Yeboah, L. Juhász, S. Coetzee, P. Mooney, A. Sarretta, & J. Anderson (Eds.). *Proceedings of the OSM Science 2023* (pp. 26–28). Zenodo.

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

[32] OSM Contributors (2023). *Vandalism and blocks in Israel*. https://community.openstreetmap.org/t/ vandalism-and-blocks-in-israel/105176

[33] Juhász, L., Mooney, P., Hochmair, H. H., & Guan, B. (20230). *ChatGPT as a mapping assistant: A novel method to enrich maps with generative AI and content derived from street-level photographs*. arXiv. https://arxiv.org/abs/2306.03204

[34]. Overture Maps Foundation (2022). *Overture Maps Foundation.* https://overturemaps.org