3-31-2022

OSM Science - The Academic Study of the OpenStreetMap Project, Data, Contributors, Community, and Applications

A. Yair Grinberger
Hebrew University of Jerusalem

Marco Minghini
Joint Research Centre, EC

Levente Juhasz
Florida International University, ljuhasz@fiu.edu

Godwin Yeboah
University of Warwick

Peter Mooney
National University of Ireland, Maynooth

Recommended Citation

Grinberger, A. Y., Marco Minghini, Levente Juhász, Godwin Yeboah, and Peter Mooney. 2022. "OSM Science—The Academic Study of the OpenStreetMap Project, Data, Contributors, Community, and Applications" ISPRS International Journal of Geo-Information 11, no. 4: 230. https://doi.org/10.3390/ijgi11040230

International Journal of
Geo-Information

MDPI

Editorial

OSM Science—The Academic Study of the OpenStreetMap Project, Data, Contributors, Community, and Applications

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

1 Department of Geography, The Hebrew University of Jerusalem, Mt. Scopus, Jerusalem 91900, Israel
2 European Commission, Joint Research Centre (JRC), 21027 Ispra, Italy; marco.minghini@ec.europa.eu
3 GIS Center, Florida International University, Miami, FL 33199, USA; ljuhasz@fiu.edu
4 Information and Digital Group, University of Warwick, Coventry CV4 7AL, UK; g.yeboah@warwick.ac.uk
5 Department of Computer Science, Maynooth University, W23 F2H6 Maynooth, Co. Kildare, Ireland; peter.mooney@mu.ie
* Correspondence: yair.grinberger@mail.huji.ac.il
† 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.

Abstract: This paper is an Editorial for the Special Issue titled "OpenStreetMap as a multidisciplinary nexus: perspectives, practices and procedures". The Special Issue is largely based on the talks presented in the 2019 and 2020 editions of the Academic Track at the State of the Map conferences. As such, it represents the most pressing and relevant issues and topics considered by the academic community in relation to OpenStreetMap (OSM)—a global project and community aimed to create and maintain a free and editable database and map of the world. In this Editorial, we survey the papers included in the Special Issue, grouping them into three research perspectives: applications of OSM for studies within other disciplines, OSM data quality, and dynamics in OSM. This survey reveals that these perspectives, while being distinct, are also interrelated. This calls for the formalization of an 'OSM science' that will provide the conceptual grounds to advance the scientific study of OSM, not as a set of individualized efforts but as a unified approach.

Keywords: OpenStreetMap; perspectives; practices; procedures

Citation: Grinberger, A.Y.; Minghini, M.; Juhász, L.; Yeboah, G.; Mooney, P. OSM Science—The Academic Study of the OpenStreetMap Project, Data, Contributors, Community, and Applications. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 230. https://doi.org/10.3390/ijgi11040230

Received: 10 March 2022
Accepted: 25 March 2022
Published: 30 March 2022

OpenStreetMap (OSM) has emerged as a global project and community operating with the objective of creating and maintaining a free and editable database and a map of the world based on the contributions of volunteer mappers [1]. With its database including almost 7.5 billion data points (nodes), contributed by approximately 1.8 million users as of March 2022 [2], it is perhaps the most accomplished example of a crowdsourced geoinformation project and of the concept of volunteered geographic information [3]. As such, it has attracted considerable attention from multiple actors, including industries, governments, and humanitarian organizations [4–6]. OSM was also widely used to address the COVID-19 pandemic response [7]. Unsurprisingly, the academic community has also shown a longstanding interest in OSM, documented in several recent review papers [8–10]. One expression of this interest is the establishment of the OSM science mailing list on 17 January 2018 [11], which is dedicated to correspondence about academic research related to OSM. To the best of our knowledge, this is the sole use of the term OSM science outside of informal conversations (an impression confirmed by a Google Scholar search of the terms "OpenStreetMap science" or "OSM science" [12]). While calling the study of OSM a new 'science' (i.e., a scientific discipline or approach) may be slightly excessive, it certainly marks an emerging field distinct with its focus on the OSM project, data, contributors, community, and applications.

This Special Issue (SI), titled "OpenStreetMap as a multidisciplinary nexus: perspectives, practices and procedures" [13], is an attempt to exemplify the richness and diversity of the field, showcasing cutting edge OSM research (the first attempt since the publication of the book "OpenStreetMap in GIScience" in 2015 [14]). It was conceived during the preparations for the State of the Map (SotM) 2019 conference, i.e., the annual global meeting of the OSM community [15]. The 2019 SotM, which took place in Heidelberg, Germany, included the second edition of an academic track dedicated to scientific studies on OSM [16]. The initial intention was that the SI will incorporate papers based on the talks included in the track. Eventually, the invitation was extended to participants of the third edition of the track at the 2020 SotM (held online due to the COVID-19 pandemic) and also to scholars who did not attend the conference. Hence, the 14 papers included in the SI do indeed represent the most pressing and relevant issues and topics in relation to OSM considered by the academic community today. As the title of the SI indicates, this is not a homogenic collection. In terms of the identity of authors, the papers vary in terms of authors' disciplines, their geographic locations (with authors located in four of the five inhabited continents), and gender. The SI is especially inclusive in this latter sense with four papers having a female as a first author and a total of eight having at least one female author (57%, which was much more than the 28% found by [17]). In terms of content, the papers vary in relation to the three P's appearing in the SI's title: perspectives (what issue is being studied), practices (what is the approach for studying the issue), and procedures (how the approach is being implemented). This heterogeneity is characteristic, in general, to OSM studies, as shown by a review of OSM-related publications included in this SI [10]. Below we survey the papers included in the SI (grouped by perspective), review the practices and procedures within each group, and discuss the meaning of OSM science emerging from this survey.

Perspective 1—Applications of OSM data: This group of papers utilize OSM data for investigating issues within the scope of specific disciplines or research areas. For example, Feldmeyer et al. [18] use machine and deep learning approaches with OSM data to develop socio-economic indicators for a number of variables, including unemployment and migration. Thus, OSM becomes a very valuable source of data and the authors conclude that OSM has "untapped potential for knowledge generation" (page 13). The usefulness of OSM for assessing living conditions and monitoring their progress over time is also exemplified by Van den Hoek and his colleagues' four-year study [19], which explores how OSM data can help to assess the Sustainable Development Goals (SDGs) defined by the United Nations. After collecting OSM data in 28 refugee and 26 nearby non-refugee settlements in Uganda, the authors used OSM tags and editing history data as input for a model which assesses 11 out of the 17 SDGs, relevant to poverty and hunger, clean water and sanitation, education, gender equality, peace and justice, and so on. The utility of OSM is perhaps most evident in the representation of areas such as refugee camps, which are not well mapped in other platforms. Hence, it is not surprising that utilizing OSM in the context of informal settlements seems to be a common practice, with Soman et al. [20] producing a worldwide detailed mapping of informal settlements by using OSM road network and building footprint data to delineate land parcels and street blocks and to flag land parcels inaccessible to their respective street blocks as informal settlements. This also helps to identify the minimal road network extension required for connecting these blocks and parcels to critical services, hence offering practical solutions for local authorities.

A common theme among several papers under this perspective perhaps represents a new and emerging practice within OSM science, i.e., that which incorporates OSM data into advanced computational models, and specifically into machine learning (ML) approaches. Another related practice is the fusion of OSM and remote sensing (RS) data to produce more accurate datasets via ML methods. For example, Brauchler and Stoffels [21] develop a methodology that combines geographic object-based image analysis (GEOBIA) with OSM data to derive accurate forest-type maps. The authors show that the method can be extended to provide similar results in other geographic areas and can be extended based on the collaborative nature of OSM to feed information back to the database and update the leaf type of existing forest patches. Similarly, Ludwig et al. [22] propose a methodology for fusing OSM and Sentinel-2 imagery data to improve the estimates of public urban green spaces by using context indicators, such as accessibility to green spaces, to counterbalance potential inconsistencies and incompleteness in the underlying data. Yet, the relationship between OSM data and RS/ML is not a one-way street. In a similar way to using OSM data for enhancing RS via ML, RS and ML can contribute to mapping processes, as Albrecht et al. [23] show in their application of artificial neural networks to vertical aerial photographs for identifying OSM data that can benefit from an update. Generated through the big geospatial data platform known as Physical Analytics Integrated Repository and Services, the output of this deep learning technique highlights areas where OSM data should be updated and recommends tags to perform such an update to OSM contributors.

Perspective 2—OSM data quality: One question that underlies all of the studies that center on applications of OSM is whether the data are good enough for a specific purpose. Sometimes this question is implicit within the work; however, at times, this is asked explicitly, as in the case of the paper by Bustamante and colleagues [24] who assess the fitness of OSM data for tourism applications by extracting tourism-related indicators (e.g., number of hotel rooms and ATMs) from OSM, finding that compared to data generated by the authoritative World Economic Forum organization, OSM data provide a fairly accurate picture of official tourism statistics. This issue of fitness for use relates to the wider issue of data quality, which is a primary thread within OSM science and a key point of interest for a second cluster of papers in this SI. Despite the extensive literature on the matter, these papers present several innovations in terms of practices and procedures. This can be witnessed in the paper by Seto et al. [25] who uniquely provide a quantitative and qualitative analysis of OSM Notes data and also uncover regional differences in the use of notes, thus demonstrating how insights regarding practices of quality management can be garnered from these data.

The two papers mentioned above follow the two most common practices in OSM data quality analysis: extrinsic analysis [24], i.e., comparing OSM data with external datasets, and intrinsic analysis [25], which develops endogenous measures that determine quality indirectly based on trends in the data (e.g., seeing saturation in the number of contributions over time as an indication for data completeness). The practice of extrinsic analysis is also followed, to a certain extent, by Yeboah et al. [26] who present an innovative experimental framework which compares data produced during a stage of remote mapping (mapping based on satellite imagery) to data produced at a later stage of on-the-ground mapping, showing that data completeness relates to morphological complexity. Madubedube et al. [27], on the other hand, present an intrinsic quality analysis using patterns identified in the data through unsupervised learning which provide insights regarding their quality, e.g., "that the behavior of a small group of active contributors makes it possible to characterize the quality of an entire datasets" (page 14) due to participation inequality. Klonner et al. [28] also follow an intrinsic approach, yet the authors focus in this paper not only on quality analysis but also on the communication of results to end-users. They present the Sketch Map Tool, which relies on OSM data to perform multiple operations required for participatory mapping, including intrinsic data quality analysis. Results are communicated to project managers through a traffic light system accompanied by a set of automatically generated comments and recommendations.

Perspective 3—Dynamics in OSM: The group of papers discussed above, and especially [26,27], show how the dynamics within data collection procedures and within the OSM community impact the project. Given this, it is not surprising that the science community also engages with these dynamics, even independently from data quality issues. An example of that in this SI is the paper by Juhász et al. [29] which fills in an important gap in the literature by exploring cartographic vandalism through a data-driven study. The paper identifies acts of carto-vandalism in OSM that originate from players of the Pokémon GO smartphone game, with the discovery that most vandals do not sustain their harmful activity in the long run and that the OSM community is getting more efficient in recognizing and fixing vandalism. Schott et al. [30] offer another account of OSM dynamics through their analytical framework which allows tracing the effects of community happenings on the behavior of mappers. By analyzing behavioral changes in terms of engagement and skill, digital locations (which regions are mapped and their relation to the mapper), and community involvement, the authors find a significant similarity between the type of the happening (i.e., remote mapathon or field mapping party) and the mapping practices followed by newcomers who continue to contribute after it. The final contribution to the SI [10] also relates to the issue of dynamics, while also illuminating a previously unexplored aspect of it—the relationship between the OSM mapping community and the OSM science community. Through a literature review and several in-depth interviews, the authors find that the way OSM is perceived and studied is related to the extent of engagement of the researcher with the mapping community. This engagement is found to be facilitated by a set of bridges (e.g., mapper colleagues, community resources) but also inhibited by barriers (communication style and channels, objectivity as a scientific value) which may limit one's understanding of OSM, hence stressing the need for triangulation, i.e., validating one's knowledge through multiple sources, when studying OSM or publishing studies about it.

This message resonates well with the overall image emerging from the survey of the three perspectives above—while each perspective is distinct, it is not independent from the others with clear interrelations existing between them: utilizing OSM data first requires an understanding of their quality (and quality may be part of the application, as showed in [28] or may be improved through applications of other data [23]); quality needs to be interpreted in light of the dynamics of the data production process through increased engagement; and understanding dynamics, e.g., through increased engagement, may lead to a refocusing of the scientific effort. Conceptualizing the study of OSM as 'OSM science' may provide utility in this context by framing OSM-related scientific studies not as a set of individualized efforts but as a unified approach with shared objectives and concerns that can be formalized through dedicated theoretical frameworks relevant even for novice and incidental OSM researchers. For this to happen, the full range of OSM science needs to be represented and understood. We hope that this Editorial and the SI will prove to be a step forward in this direction.

We wish to end this Editorial by thanking the participants of State of the Map's Academic Track across the years, to the authors of all papers considered for this SI (including those eventually not included in it), to the reviewers participating in the evaluation of these papers, and to the *ISPRS International Journal of Geo-Information* for providing the stage and space for this SI. Finally, we take this opportunity to express our deepest appreciation and gratitude to the OSM community for their volunteer efforts to the provision of an accurate and complete representation of our world, and for promoting the value of open data.

The Editors,

Yair Grinberger, Marco Minghini, Levente Juhász, Godwin Yeboah, and Peter Mooney

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. About OpenStreetMap. Available online: https://wiki.openstreetmap.org/wiki/About_OpenStreetMap (accessed on 6 March 2022).
2. Stats. Available online: https://wiki.openstreetmap.org/wiki/Stats (accessed on 6 March 2022).
3. Goodchild, M.F. Citizens as Sensors: The World of Volunteered Geography. *GeoJournal* **2007**, *69*, 211–221. [CrossRef]
4. Anderson, J.; Sarkar, D.; Palen, L. Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 232. [CrossRef]
5. Herfort, B.; Lautenbach, S.; Porto de Albuquerque, J.; Anderson, J.; Zipf, A. The Evolution of Humanitarian Mapping within the OpenStreetMap Community. *Sci. Rep.* **2021**, *11*, 3037. [CrossRef] [PubMed]
6. Sarretta, A.; Minghini, M. Towards the Integration of Authoritative and OpenStreetMap Geospatial Datasets in Support of the European Strategy for Data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *XLVI-4/W2-2021*, 159–166. [CrossRef]
7. Mooney, P.; Grinberger, A.Y.; Minghini, M.; Coetzee, S.; Juhász, L.; Yeboah, G. OpenStreetMap Data Use Cases during the Early Months of the COVID-19 Pandemic. In *COVID-19 Pandemic, Geospatial Information, and Community Resilience: Global Applications and Lessons*; CRC Group: Boca Raton, FL, USA, 2021; pp. 171–186.

8. Kaur, J.; Singh, J.; Sehra, S.S.; Rai, H.S. Systematic Literature Review of Data Quality Within OpenStreetMap. In Proceedings of the 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS), Jammu, India, 11–12 December 2017; pp. 177–182.
9. Sehra, S.S.; Singh, J.; Rai, H.S. Using Latent Semantic Analysis to Identify Research Trends in OpenStreetMap. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 195. [CrossRef]
10. Grinberger, A.Y.; Minghini, M.; Yeboah, G.; Juhász, L.; Mooney, P. Bridges and Barriers: An Exploration of Engagements of the Research Community with the OpenStreetMap Community. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 54. [CrossRef]
11. Mooney, P. [OSM-Science] Getting Started on Science at OSM. Available online: https://lists.openstreetmap.org/pipermail/science/2018-January/thread.html (accessed on 23 January 2022).
12. Google Scholar Query: "OpenStreetMap Science" | "OSM Science". Available online: https://scholar.google.com/scholar?hl=iw&as_sdt=0%2C5&q=%22OpenStreetMap+science%22+%7C+%22OSM+science%22&btnG= (accessed on 23 January 2022).
13. Special Issue "OpenStretMap as A Multi-Disciplinary Nexus: Perspectives, Practices and Procedures". Available online: https://www.mdpi.com/journal/ijgi/special_issues/OpenStreetMap (accessed on 6 March 2022).
14. Arsanjani, J.J.; Zipf, A.; Mooney, P.; Helbich, M. *OpenStreetMap in GIScience. Lecture Notes in Geoinformation and Cartography*; Springer International Publishing: Cham, Switzerland, 2015; ISBN 978-3-319-14279-1.
15. State of the Map. 2019. Available online: https://2019.stateofthemap.org/ (accessed on 23 January 2022).
16. Academic Programme. Available online: https://2019.stateofthemap.org/academic_programme/ (accessed on 23 January 2022).
17. Pierson, E. In Science, It Matters That Women Come Last. Available online: https://fivethirtyeight.com/features/in-science-it-matters-that-women-come-last/#fn-1 (accessed on 6 March 2022).
18. Feldmeyer, D.; Meisch, C.; Sauter, H.; Birkmann, J. Using OpenStreetMap Data and Machine Learning to Generate Socio-Economic Indicators. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 498. [CrossRef]
19. Van Den Hoek, J.; Friedrich, H.K.; Ballasiotes, A.; Peters, L.E.R.; Wrathall, D. Development after Displacement: Evaluating the Utility of OpenStreetMap Data for Monitoring Sustainable Development Goal Progress in Refugee Settlements. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 153. [CrossRef]
20. Soman, S.; Beukes, A.; Nederhood, C.; Marchio, N.; Bettencourt, L. Worldwide Detection of Informal Settlements via Topological Analysis of Crowdsourced Digital Maps. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 685. [CrossRef]
21. Brauchler, M.; Stoffels, J. Leveraging OSM and GEOBIA to Create and Update Forest Type Maps. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 499. [CrossRef]
22. Ludwig, C.; Hecht, R.; Lautenbach, S.; Schorcht, M.; Zipf, A. Mapping Public Urban Green Spaces Based on OpenStreetMap and Sentinel-2 Imagery Using Belief Functions. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 251. [CrossRef]
23. Albrecht, C.M.; Zhang, R.; Cui, X.; Freitag, M.; Hamann, H.F.; Klein, L.J.; Finkler, U.; Marianno, F.; Schmude, J.; Bobroff, N.; et al. Change Detection from Remote Sensing to Guide OpenStreetMap Labeling. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 427. [CrossRef]
24. Bustamante, A.; Sebastia, L.; Onaindia, E. On the Representativeness of OpenStreetMap for the Evaluation of Country Tourism Competitiveness. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 301. [CrossRef]
25. Seto, T.; Kanasugi, H.; Nishimura, Y. Quality Verification of Volunteered Geographic Information Using OSM Notes Data in a Global Context. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 372. [CrossRef]
26. Yeboah, G.; Porto de Albuquerque, J.; Troilo, R.; Tregonning, G.; Perera, S.; Ahmed, S.A.K.S.; Ajisola, M.; Alam, O.; Aujla, N.; Azam, S.I.; et al. Analysis of OpenStreetMap Data Quality at Different Stages of a Participatory Mapping Process: Evidence from Slums in Africa and Asia. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 265. [CrossRef]
27. Madubedube, A.; Coetzee, S.; Rautenbach, V. A Contributor-Focused Intrinsic Quality Assessment of OpenStreetMap in Mozambique Using Unsupervised Machine Learning. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 156. [CrossRef]
28. Klonner, C.; Hartmann, M.; Dischl, R.; Djami, L.; Anderson, L.; Raifer, M.; Lima-Silva, F.; Castro Degrossi, L.; Zipf, A.; Porto de Albuquerque, J. The Sketch Map Tool Facilitates the Assessment of OpenStreetMap Data for Participatory Mapping. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 130. [CrossRef]
29. Juhász, L.; Novack, T.; Hochmair, H.; Qiao, S. Cartographic Vandalism in the Era of Location-Based Games—The Case of OpenStreetMap and Pokémon GO. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 197. [CrossRef]
30. Schott, M.; Grinberger, A.Y.; Lautenbach, S.; Zipf, A. The Impact of Community Happenings in OpenStreetMap—Establishing a Framework for Online Community Member Activity Analyses. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 164. [CrossRef]