Proceedings of the Academic Track

Editors:

Marco Minghini
A. Yair Grinberger
Peter Mooney
Levente Juhász
Godwin Yeboah

Introduction

State of the Map is the annual meeting of the international OpenStreetMap community. In 2019 it was held in Heidelberg, Germany on September 21-23. For the second year, State of the Map featured a full day of academic talks. Building upon the motto of SotM 2019 in “Bridging the Map” the Academic Track session was aimed to provide the bridge to join together the experience, understanding, ideas, concepts and skills from different groups of researchers, academics and scientists from around the world. In particular, the Academic Track session was meant to build this bridge that connects members of the OpenStreetMap community and the academic community by providing an open passage for exchange of ideas, communication and opportunities for increased collaboration.

These proceedings include 14 abstracts accepted as oral presentations and 6 abstracts presented as posters. Contributions were received from different academic fields, for example geography, remote sensing, computer and information sciences, geomatics, GIScience, the humanities and social sciences, and even from industry actors. We are particularly delighted to have included abstracts from both experienced researchers and students. Overall, it is our hope that these proceedings accurately showcase the ongoing innovation and maturity of scientific investigations and research into OpenStreetMap, showing how it as a research object converges multiple research areas together. Our aim is to show how the sum total of investigations of issues like Volunteered Geographic Information, geo-information, and geo-digital processes and representation shed light on the relations between crowds, real-world applications, technological developments, and scientific research.

The Scientific Committee would like to thank the following people for their support to the Academic Track: Mrs. Christine Karch, Mr. Michael Reichert, and Dr. Martin Raifer. Also, the Scientific Committee expresses gratitude to the OpenStreetMap Foundation for inviting academic researchers into the State of the Map 2019 conference and community.

Contents

Bridging the Map? Exploring Interactions between the Academic and Mapping Communities in OpenStreetMap

A. Yair Grinberger[1,2]*, Marco Minghini[3], Levente Juhász[4], Peter Mooney[5] and Godwin Yeboah[6]

[1] Department of Geography, The Hebrew University of Jerusalem, Israel; yair.grinberger@mail.huji.ac.il
[2] GIScience Research Group, Heidelberg University, Heidelberg, Germany;
[3] European Commission, Joint Research Centre (JRC), Ispra, Italy; marco.minghini@ec.europa.eu
[4] GIS Center, Florida International University, Miami, FL, United States; ljuhasz@fiu.edu
[5] Department of Computer Science, Maynooth University, Maynooth, Ireland; peter.mooney@mu.ie
[6] Institute for Global Sustainable Development, School of Cross-faculty Studies, University of Warwick, Coventry, United Kingdom; g.yeboah@warwick.ac.uk

* 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.

OpenStreetMap (OSM) can be conceptualized in a multitude of ways: it may be seen as a database, as a platform, as a concept, as a community (or collection of communities), as a social practice, etc. The academic research on OpenStreetMap adopts and utilizes these different conceptualizations, creating various forms of inquiry. For example, quality-related inquiries can be linked to the data/platform perspectives [1, 2], contributor behaviors are analyzed quantitatively and qualitatively from a more behavioral perspective [3, 4], and social understandings of OSM are utilized in inquiries into the institutional and community dimensions of the project [5, 6]. Indicative of a more general issue in the relations between geo-information and socio-cultural contexts [7], these readings of OSM do not represent absolute truths, but rather they emerge from the specific personal, professional, and socio-cultural backgrounds of OSM researchers (OSM-R). Furthermore, they hold the potential to create an effect on the world and specifically on OSM and its communities. However, the extent and nature of these relations in OSM-R, and specifically relations between research and the OSM community (OSM-C) have not received much academic attention yet. This is despite such interactions existing, e.g. when research outputs are presented to the community, when OSM contributors (OSMappers) become researchers themselves and vice versa, or on other occasions.

Efforts to establish and strengthen the interaction between OSM-R and OSM-C have already resulted into significant outputs, e.g. the creation of a dedicated 'OSM science' mailing list and the stable inclusion of an Academic Track into the annual State of the Map conference. In this study, we make a step further in the exploration of this issue, with the objective of not only better understanding these interactions but also formalizing an agenda for future OSM-R endeavors. Specifically, we look at the interactions between OSMappers

Grinberger et al. (2019). Bridging the Map? Exploring Interactions between the Academic and Mapping Communities 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, 1-2. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3408639

and research communities, analyzing how the two affect each other, what are the implications of these interactions for both the researchers and the community, and how could these be changed to enhance relations and make them more productive ones. While this issue can be studied from the perspectives of both OSM-R and OSM-C, we focus on an initial exploration of the former.

For this purpose, we employ two techniques. First, we review OSM-R publications from recent years (2016-2019) and, in addition to classifying them according to the researchers' background discipline and the topic, consider what type of conceptualizations of OSM are employed there, and whether and how interactions with the OSM-C are considered explicitly. We use this analysis to make an initial assessment of the state of the issue in the field and identify how specific topics/backgrounds affect the ways in which OSM is conceptualized in research. Second, we collect detailed records of experiences of OSM-R/OSM-C interactions via the self-reflections of the authors and interviews with colleagues. While far from representative of the entire field, these allow a deeper observation of the causal processes that lead to the adoption of certain perspectives and to the development (or lack) of OSM-R/OSM-C interactions. In such a way, we gain insights into how researchers that are also mappers manage their different community roles and sets of objectives, when interactions (if any) happen, what their nature is, who initiates them, who dominates them, and why these came to be that way. Furthermore, these reflections allow speculation on how things could have been done differently, which opportunities were missed, and what possibilities exist. Thus, the combination of a view of current research status with an understanding of processes and forward-looking thinking allow us to point towards possible steps and procedures OSM-R could consider in order to create an impact on OSM-C and to enhance research via an understanding of OSM as a community.

References

[1] Haklay, M. (2010). How good is volunteered geographical information? A comparative analysis of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, *37*(4), 682-703.

[2] Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, *18*(6), 877-895.

[3] Lin, Y. W. (2011). A qualitative inquiry into OpenStreetMap making. *New Review of Hypermedia and Multimedia*, *17*(1), 53-71.

[4] Bégin, D., Devillers, R., & Roche, S. (2017). Contributors' withdrawal from online collaborative communities: The case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *6*, 340.

[5] 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.

[6] Fast, V., & Rinner, C. (2014). A systems perspective on volunteered geographic information. *ISPRS International Journal of Geo-Information*, *3*(4), 1278-1292.

[7] Chrisman, N. (2005). Full circle: More than just the social implications of GIS. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *40*(4), 23-35.

Exploring the Effects of Pokémon Go Vandalism on OpenStreetMap

Levente Juhász[1,*], Hartwig Hochmair[2], Sen Qiao[1] and Tessio Novack[3]

[1] GIS Center, Florida International University, Miami, FL, United States; ljuhasz@fiu.edu, sqiao@fiu.edu
[2] Geomatics Program, University of Florida, Fort Lauderdale, FL, United States; hhhochmair@ufl.edu
[3] GIScience Research Group, Heidelberg University, Heidelberg, Germany; novack@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.

The utilization of OpenStreetMap (OSM) data by mainstream tech companies has been on the rise in recent years. Two prominent examples are Snapchat and Pokémon Go that became OSM data consumers in 2017. Snapchat reports 190 million daily active users in 2019 [1]. Pokémon Go was used by 28.5 million users daily during its peak popularity in 2016 and it still managed to engage more than 10 million users monthly in 2018 [2]. The large user base of these applications puts OSM in an unprecedented spotlight which can be considered a huge success for the project. On the other hand, increased attention comes with undesirable side effects. Acts of vandalism [3] manifested in the data no longer stay within the OSM community but will be visible to a worldwide audience. This increased visibility of errors caused by malicious actions (e.g. fake place names, fictive data) can potentially undermine the reputation of the OSM project. In August 2018, a case of anti-semitic vandalism surfaced on Snapchat's online maps [4] and also made it to various mainstream media outlets, such as the BBC, Time or The New York Times. Another type of vandalism can be observed in connection with Pokémon Go, where users modify the underlying OSM data by adding fictional map features (e.g. parks, footpaths and lakes) to gain benefits in the game [5].

OSM's vulnerability to vandalism is often considered one of its drawbacks directly related to data quality. Despite this and other negative effects on the OSM project, carto-vandalism [6] has only been addressed sporadically in the literature. One study identified motivations behind such actions [7], while some other studies characterized different types of vandalism based on investigations of community forums and mailing lists [6] and documented cases of vandalism [8]. According to Linus's law, the collaborative nature of OSM ensures that vandalism will be discovered and corrected [9]. However, it is unreasonable to expect that all harmful contributions will be found by community [10], therefore, automatic detection of vandalism with rule-based methods is of interest [8, 11]. The OSM community also developed a set of tools to battle vandalism.

Using Pokémon Go as an example, this study focuses on the nature and life-cycle of harmful edits with an emphasis on the OSM community's response. Based on OSM changeset comments and discussions, the study first identifies Pokémon Go related

Juhász et al (2019). Exploring the Effects of Pokémon Go Vandalism on 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, 3-4. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3386533

vandalism together with changesets that fixed them or reverted them. This duality allows to study not only the act of vandalism itself, but also the community's response. By analyzing Pokémon Go vandalism changesets this study describes in detail what fake information have been added to the map. A better understanding of this will allow to develop more targeted rules for vandalism detection systems.

It is important to note that not all Pokémon Go players vandalize OSM. It is well known that several Pokémon Go players are also valuable members of the OSM community. Apart from gaining benefits in the game, an alternative explanation for Pokémon Go vandalism might be that those users are not aware of the purpose of the OSM project, therefore they do not even realize the implications of adding fake data. Pokémon Go players can be considered a large pool of potential OSM community members if they do not vandalize the map. It was observed that instead of just fixing harmful edits several experienced members of the OSM community reach out to mappers who initially added fake data (e.g. through changeset discussions). Therefore, this study seeks empirical evidence of initial “vandals” converted to be constructive OSM contributors due to the interactions with other mappers. A better understanding of what communication techniques worked would help utilizing OSM's increased visibility to engage and retain more contributors.

Vandalism directly affects data quality, therefore this study aims to provide a first description of the life-cycle of carto-vandalism analyzing a large pool of events and considering both spatial and temporal constraints. Our initial data analysis identified more than 1,500 changesets that reverted harmful Pokémon Go edits. These revert changesets fixed more than 4,000 changesets that can be considered vandalism.

References

[1] Snap, Inc. (2019). First Quarter 2019 Financial Results. Retrieved from https://bit.ly/2YAzv70

[2] Statista (2019). Most popular mobile gaming related apps in the United States as of March 2019. Retrieved from https://bit.ly/2VJS3EO

[3] OpenStreetMap Wiki (2019). Vandalism. Retrieved from https://wiki.openstreetmap.org/wiki/Vandalism

[4] OSM Foundation (2018). OSM condemns recent anti-semitic vandalism. Retrieved from https://bit.ly/2HHPRno

[5] Huffington Post (2018). The Strange Story Of How Pokemon Go Put Lakes And Parks In People's Homes. Retrieved from https://bit.ly/2EjFSnD

[6] Ballatore, A. (2014). Defacing the map: Cartographic vandalism in the digital commons. *The Cartographic Journal, 51*(3), 214-224.

[7] Coleman, D., Georgiadou, Y., & Labonte, J. (2009). Volunteered geographic information: The nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research, 4*(1), 332-358.

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

[9] Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning and design, 37*(4), 682-703.

[10] Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial statistics, 1*, 110-120.

[11] Truong, Q. T., Touya, G., & De Runz, C. (2018). Towards Vandalism Detection in OpenStreetMap Through a Data Driven Approach. In *Proceedings of the 10th International Conference on Geographic Information Science (GIScience 2018)*, 61:1-61:7.

A novel application of models of species abundance to better understand OpenStreetMap community structure and interactions

Peter Mooney[1,*]

[1] Department of Computer Science, Maynooth University, Maynooth, Ireland; peter.mooney@mu.ie

* 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.

Society is built around people coming together into social groups called communities. Communities identify different groups and very often the bond within communities is a set of shared goals and the division/sharing of labour, skills and other resources. The feeling of contributing positively to our own communities is one of the most fundamental feelings of satisfaction in life [1]. All of these traits are visible within the millions of contributors to OSM which form the OSM Community. Attempts to understand how the OSM community works have appeared in the academic literature, such as in [2]. There is a curiosity and fascination about the OSM community given the global extent of OSM: crossing cultures, geographical boundaries and languages; the altruistic nature of its members; and OSM's success as a primarily Internet-based community. In this paper we argue that the model of community required for OSM is more nuanced that many of the current quantitative approaches.

Of many suggested contribution models, OSM has been shown to loosely adhere to the 90-9-1 rule of contribution. In [3] the authors highlight that about 90% of the members of community-based projects only consume the collected information, while 9% occasionally contribute and only 1% demonstrate a very active pattern of contribution. As argued in [4], "characterizing VGI data requires understanding contributors' behaviour and many typologies of contributors are proposed in an attempt to link VGI contributors with the nature of the data they provide". In [5] the authors identify different phases of contributor life cycle from a temporal perspective as a contributor's lifespan is a 'university metric'. In a more computationally complex approach [6], a multigraph approach with data mining is developed to characterise individuals and identify behavioural groups. We consider a very novel approach to community identification and understanding by borrowing concepts and methodologies from theories and models of species abundance to the individual contributors of the OSM community. This is a novel approach for VGI but well established within Ecological Sciences.

Mooney (2019). A novel application of models of species abundance to better understand OpenStreetMap Community structure and interactions
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, 5-6. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3386559

Hughes [7] argues that “in animal and plant communities most of the individuals belong to a small number of abundant species, whereas most of the species are represented by a small number of individuals”. In OSM we see that most individual contributors make a small number of contributions. However, from the global OSM community, a small number of species (groupings) are represented by a small number of contributors (for example very high frequency contributors). We use the OSM Planet History data for a number of selected regions to consider OSM contribution history in those regions. We then develop and apply the Community Level Models (CLMs) from authors in [8] as well as others. We define different types of OSM community member species. For example, we may create species which are differentiated by the number of OSM relations they have created/edited. More sophisticated species can then be developed and integrated into different models. CLMs allow the creation of species co-occurrence matrices to environmental variables (such as quantity of edits, types of tagging, etc.) allowing prediction of the community structure and the distributions of species. CLMS can predict species distributions and changes in the community composition more accurately than other methodologies [8]. Assuming OSM contributors exist in isolation and do not influence one another's editing potentially limits understanding of contribution patterns [9]. This innovative approach potentially offers better understanding of interactions between contributors: for example, between experienced contributors and new entrants or editing interactions during events such as mapping parties. Comparison and analysis of CLMs applied to different countries and regions is also of great value.

References

[1] Proctor, N. (2013). Crowdsourcing—An Introduction: From Public Goods to Public Good. *Curator: The Museum Journal, 56(*1), 105-106.

[2] 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.

[3] Carron-Arthur, B., Cunningham, J. A., & Griffiths, K. M. (2014). Describing the distribution of engagement in an Internet support group by post frequency: A comparison of the 90-9-1 Principle and Zipf's Law. *Internet Interventions*, *1*(4), 165-168.

[4] Bégin, D., Devillers, R., & Roche, S. (2017). Contributors’ Withdrawal from Online Collaborative Communities: The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *6*(11), 340.

[5] Bégin, D., Devillers, R., & Roche, S. (2017). Contributors’ enrollment in collaborative online communities: the case of OpenStreetMap. *Geo-spatial Information Science*, *20*(3), 282-295.

[6] Truong, Q. T., De Runz, C., & Touya, G. (2019). Analysis of collaboration networks in OpenStreetMap through weighted social multigraph mining. *International Journal of Geographical Information Science*, *33*(8), 1651-1682.

[7] Hughes, R. G. (1986). Theories and models of species abundance. *The American Naturalist*, *128*(6), 879-899.

[8] Maguire, K. C., Nieto-Lugilde, D., Blois, J. L., Fitzpatrick, M. C., Williams, J. W., Ferrier, S., & Lorenz, D. J. (2016). Controlled comparison of species-and community-level models across novel climates and communities. In *Proceedings of the Royal Society B: Biological Sciences*, *283*(1826), 20152817.

[9] Mooney, P., & Corcoran, P. (2014). Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. *Transactions in GIS*, *18*(5), 633-659.

Analysis of OpenStreetMap data quality at different stages of a participatory mapping process: Evidence from informal urban settings

Godwin Yeboah[1,*], Rafael Troilo[2], Vangelis Pitidis[1] and João Porto de Albuquerque[1]

[1] Institute for Global Sustainable Development, University of Warwick, Coventry, United Kingdom; g.yeboah@warwick.ac.uk, v.pitidis@warwick.ac.uk, j.porto@warwick.ac.uk
[2] GIScience Research Group, Heidelberg University, Heidelberg, Germany; rafael.troilo@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.

Globally, the lack of detailed quality spatial data of informal urban settings, such as slums, is increasingly becoming a concern to both researchers and development agencies [1, 2]. One potential for making spatial data available is through Volunteered Geographic Information which is opening up new possibilities of data production in recent years and facilitating the emergence of several initiatives aimed at "putting the most vulnerable people on the map" [3]. The increasing availability of volunteered and crowdsourced geographic information, in particular OpenStreetMap (OSM), has led to a plethora of scientific studies with emphasis on evaluating the quality of the OSM data. The quality assessment results are usually presented in the form of tables, diagrams, maps and statistics per given area [4, 5]. Some recent studies have examined OSM data quality without using any external data; the so called intrinsic approach [4]. In contrast to intrinsic approach, other studies commonly used what is referred to as the extrinsic approach where the OSM data is compared with external datasets such as the UK Ordnance Survey data [6]. In both approaches, the data production processes are often not completely transparent to researchers therefore limiting possibilities for systematic data quality analysis of the processes leading to OSM update.

This presentation examines OSM data quality at different stages of a participatory mapping process developed as part of an ongoing multi-country research project focused on understanding inequalities in healthcare access of slum residents in the Global South [7]. The following research questions are addressed: (1) What is the level of spatial data quality one can expect at different stages of the mapping process leading to final update of the OpenStreetMap database? (2) What are the factors influencing quality? The OSM project provides the entire history containing all edits made by community of mappers. Examining the evolution of the OSM history data could lead to new insights about its quality and

Yeboah et al. (2019). Analysis of OpenStreetMap data quality at different stages of a participatory mapping process: Evidence from informal urban settings
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, 7-8. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3386563

completeness. Due to the complexity, heterogeneity and size of the data, its analysis becomes a challenging endeavour. Therefore, our exploratory method applies a recently developed data analytics framework for spatio-temporal analysis, the ohsome platform [8], which allows us to analyse our study areas at different points in time. The historical data sets analysed are derived from the following stages in the participatory mapping process: before online mapping as stage 1, after online mapping and validation but before ground-truthing as stage 2, and, after ground-truthing as stage 3. The before-and-after estimates at each mapping stage are discussed together with lessons learnt. We thus present initial results from a spatial data quality assessment of the mapping process stages used to map our study areas and update the OSM database. We define completeness differential as the difference in estimates of OSM features at timestamps K and X during the participatory mapping process, where K is the feature estimate at the end of stage 2 and X is the feature estimate at the end of stage 3 or in percentage terms X divided by K times hundred.

Findings show 82% for Azam Basti (Pakistan), 99% for Sasa (Nigeria) and 63% for Korogocho (Kenya). Completeness differentials of buildings show varying differences with stage 3 showing relatively lower estimates than stage 2 for areas with unique rooftop architecture and dense structures. This means that using OSM data after stage 2 in such areas is not desirable; especially for actions such as survey research. However, for areas like Sasa with less dense structures, completeness differential is small and therefore data from stage 2 could be used for both survey research and disaster response operations with less risk. Rooftop architecture has an impact on quality of building completeness, particularly in the case of Azam Basti where the rooftop architecture representation in the satellite imagery made the interpretation of the building footprints difficult. This experience rendered stage 2 data for Azam Basti less useful during field work. Our participatory mapping process is reproducible given that the same mapping workflow and open source technologies were used across all the study sites with different local teams. However, the process still requires technical know-how and future work should focus on optimizing the integration of the tools used to make it easier to implement for survey research.

References

[1] Hachmann, S., Jokar Arsanjani, J., & Vaz, E. (2018). Spatial data for slum upgrading: Volunteered Geographic Information and the role of citizen science. *Habitat International*, *72*, 18-26.

[2] Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from space—15 years of slum mapping using remote sensing. *Remote Sensing*, *8*(6), 455.

[3] Missing Maps (2019). Putting the World's Vulnerable People on the Map. Retrieved from http://www.missingmaps.org.

[4] Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, *18*(6), 877-895.

[5] Sehra, S., Singh, J., & Rai, H. (2017). Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. *Future Internet*, *9*(2), 15.

[6] Haklay, M. (2010). How good is volunteered geographical information? A comparative analysis of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design, 37*(4), 682-703.

[7] National Institute of Health Research (NIHR) Global Health Research Unit on Improving Health in Slums project (2017). Retrieved from https://warwick.ac.uk/ fac/sci/med/about/centres/cahrd/slums

[8] Heidelberg Institute for Geoinformation Technology (2019). Big Spatial Data Analytics. Retrieved from https://heigit.org/big-spatial-data-analytics-en

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

A. Yair Grinberger[1,2]*, Moritz Schott[2], Martin Raifer[2], Rafael Troilo[2] and Alexander Zipf[2]

[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

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., & Rehrl, 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.

Development after displacement: Using OSM data to measure Sustainable Development Goal indicators at informal settlements

Jamon Van Den Hoek[1,*], Hannah Friedrich[1], Anna Ballasiotes[1] and David Wrathall[1]

[1] College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, United States; vandenhj@oregonstate.edu, friedrih@oregonstate.edu, ballasia@oregonstate.edu, wrathald@oregonstate.edu

* 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.

In 2015, the United Nations introduced 17 Sustainable Development Goals (SDGs) and 169 associated targets as part of its 2030 Agenda. One hundred and ninety three countries declared their commitment to "leave no one behind" in the shared pursuit of SDGs, yet 250 million people around the globe are estimated to be missing from SDG progress assessments. This presentation focuses on one category of these "Missing Millions": refugees and internally displaced people (IDPs) living in informal settlements around the globe. Over two-thirds of the global refugee population lives in a "protracted refugee situation" where basic rights and access to economic and social services remain unfulfilled after years and sometimes generations in exile [1]. However, in part due to the lack of reliable, open information on the locations of informal settlements [2], refugees and IDPs have been systematically excluded from national censuses, representative surveys, and global settlement and population data sets [3]. Given that similar conditions may contribute to partial or biased SDG assessments at refugee and IDP settlements, the goal of this work is to develop a geographically consistent and open schema of SDG indicators relevant for the Missing Millions living in informal settlements.

Though there has been notable progress towards developing Earth Observation (EO) based SDG indicators [4], many amenities or public services such as a school, health center, or sites of potable drinking are not readily identified with satellite imagery. While OpenStreetMap (OSM) data excel at capturing site-specific features and have nominally global coverage, their use in evaluating SDG indicators has remained limited [5]. In examining the value of OSM data to support an SDG monitoring approach that is explicitly inclusive of these highly vulnerable populations, we considered the relatively data rich case study of across 23 settlements in Uganda. We incorporated UNHCR refugee settlement location and boundary data collected by Uganda Humanitarian OpenStreetMap Team, queried all OSM features within or nearby refugee settlements as well as non-refugee settlements, and identified all OSM feature tags. We culled this collection of OSM tags to those with direct relevance for SDG 3 (health and well-being), SDG 4 (inclusive education),

Van Den Hoek et al. (2019). Development after Displacement: Using OSM data to measure SDG indicators at informal settlements 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, 11-12. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387675

SDG 6 (clean water, sanitation), and SDG 11 (sustainable communities). We then measured the spatial proximity, thematic variability, date of appearance, and version history for each OSM tag at Ugandan refugee settlements, and assessed the presence and relevance of these tags across a broader sample of refugee settlements in 22 other African countries. We found that the majority of SDG-relevant tags are related to the amenity key and paired with values related to clinics, schools, toilets, and drinking water. The geographic distribution of these key-value pairs is not consistent across refugee settlements: many settlements lack any SDG-relevant OSM data while others may only contain data for a given SDG.

This variability speaks to the complex and varied history of OSM data collection/creation at refugee settlements, the challenges of retroactively including OSM data collected over many years and various campaigns within a single SDG assessment schema, and the difficulty of separating a true absence of SDG indicators from a false absence due to an omission of OSM data. Indeed, many data collection biases that contribute to the variability in OSM tags within and between informal settlements challenge the development of a globally relevant schema for using OSM data to assess SDG indicators at refugee settlements. For example, the presence of SDG-relevant OSM features at some settlements may reflect past humanitarian or development mapping, while other informal settlements with less attention in OSM may lack any SDG-relevant OSM features. The absence of SDG-relevant OSM features across so many informal settlements highlights the need for a rigorous assessment of potential bias in settlement-level OSM data availability as well as consistency in OSM metadata attribution to support linkages to SDG indicators consistent in time and space. Further research to characterize the conditions and consequences of using OSM data for SDG assessment may help refine OSM on-the-ground and online volunteer campaigns, guide a field-based assessment of the SDG schema, inform global end users of OSM data in settlement-level SDGs assessments, and bridge the wide gap between EO- and OSM-derived SDG indicators at informal settlements.

References

[1] Milner, J. (2014). Protracted refugee situations. In *The Oxford handbook of refugee and forced migration studies*, 151-162.

[2] Chakraborty, A., Wilson, B., Sarraf, S., & Jana, A. (2015). Open data for informal settlements: Toward a user's guide for urban managers and planners. *Journal of Urban Management*, *4*(2), 74-91.

[3] Carr-Hill, R. (2013). Missing millions and measuring development progress. *World Development*, *46*, 30-44.

[4] Anderson, K., Ryan, B., Sonntag, W., Kavvada, A., & Friedl, L. (2017). Earth observation in service of the 2030 Agenda for Sustainable Development. *Geo-spatial Information Science*, *20*(2), 77-96.

[5] MacFeely, S. (2019). The big (data) bang: Opportunities and challenges for compiling SDG indicators. *Global Policy*, *10*, 121-133.

Intrinsic assessment of OpenStreetMap contribution patterns through Exploratory Spatial Data Analysis

Marco Minghini[1,*], Daniele Oxoli[2], Francesco Frassinelli[3] and Maria Antonia Brovelli[2]

[1] European Commission, Joint Research Centre (JRC), Ispra, Italy; marco.minghini@ec.europa.eu
[2] Department of Civil and Environmental Engineering, Politecnico di Milano, Milan, Italy; daniele.oxoli@polimi.it, maria.brovelli@polimi.it
[3] Norsk institutt for naturforskning (NINA), Trondheim, Norway; francesco.frassinelli@nina.no

* 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.

Compared to traditional geospatial data sources, a major advantage of OpenStreetMap (OSM) is the availability of its full history. In literature, OSM history has been exploited for a number of purposes. The most frequent is intrinsic quality assessment, which – in contrast to extrinsic assessment, where OSM quality is evaluated through comparison against a reference dataset – estimates OSM quality by only looking at its temporal evolution. OSM history has been also explored to gain insights into the project's contribution patterns, e.g. history and profiling of contributors; origin, amount, nature and frequency of edits; spatio-temporal evolution of the whole OSM database – or parts thereof, such as road networks and buildings – in specific areas, or after specific events like natural disasters; and spatial analysis of contributor and contribution patterns.

This work fits into the context of OSM intrinsic assessment by proposing a statistical approach based on Exploratory Spatial Data Analysis, and in particular spatial association [1], aimed at uncovering underlying history-based patterns of OSM data. More in detail, spatial association is investigated in both the univariate and multivariate contexts, i.e. in the cases – respectively – when one variable and multiple variables (together) are examined. The univariate analysis is performed using the Local Moran's I indicator, which provides a robust classification method to detect statistically significant patterns (compared to the hypothesis of randomness) and defines the spatial association type at each location in the dataset [2]. The association type reflects the local characteristics of the variable at each location and its surroundings. Hence it allows detecting clusters, i.e. local patterns of similar (either high or low) values, as well as outliers, i.e. local patterns of dissimilar values (either low values surrounded by high values or viceversa). Instead, the multivariate Geary's c indicator is employed to detect local association patterns resulting from the joint spatial interaction of two or more variables [3]. A multivariate pattern classification comparable to the one of the univariate case is achieved through a novel classification method developed by the authors [4]. This consists of a comparison of local and global centrality measures

Minghini et al. (2019). Intrinsic assessment of OpenStreetMap contribution patterns through Exploratory Spatial Data Analysis 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, 13-14. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387683

(means and medians) for the computed distribution of the multivariate Geary's c, to produce classification maps of clusters and outliers.

The analysis is performed on Milan Province (Northern Italy), counting a population of more than 3 million inhabitants on a surface of about 1.500 km². This area is sampled using a regular hexagonal grid with side of the hexagon equal to 1000 m, producing a total of 684 cells. The analysis is focused on the history of OSM nodes only, with the following hypotheses: only nodes with at least one tag are considered; a new version of a node is counted only when there is a change in tags (not in geometry); only the nodes which currently exist in the OSM database are considered. With this in mind, for each grid cell a number of history-based variables (mostly derived from literature) are computed: total number of different contributors who have edited OSM nodes; average number of different contributors who have edited each OSM node; average date of creation of the OSM nodes; average date of last edit of the OSM nodes; average number of versions of the OSM nodes; average frequency of update of the OSM nodes. These values are derived from the processing of the OSM Full History Planet file (downloaded in May 2019) and its conversion into a SpatiaLite database after an intersection with the study area, followed by the computation of the variables for each grid cell.

The univariate analysis, performed using the QGIS Hotspot Analysis plugin developed by the authors [5], highlights different spatial associations for the different variables. While some of them (such as total and average number of contributors and average number of versions) clearly show clusters of high values in correspondence of the most urbanized areas and clusters of low values in the non-urban peripheral areas, spatial association patterns are more heterogeneous for other variables such as the average update frequency. Multivariate analyses are then performed to detect the spatial patterns derived from the joint interaction between two and more of the variables considered. Despite each variable has its own spatial pattern when taken alone, their combination (especially when adding more and more variables) highlights not only high and low-value clusters in urban and non-urban areas, but also other interesting clusters and outliers. These unveil peculiar contribution patterns resulting from active local contributors, data imports and mapping parties, and highlight areas where OSM development might need some improvement.

Despite preliminary, the methodology – which, to the authors' knowledge, has been never adopted before in OSM-related research – looks extremely promising to process the complexity of OSM history and transform it into understandable, statistical-based indicators which can shed more light on the intricate phenomenon of OSM local development.

References

[1] Unwin, A., & Unwin, D. (1998). Exploratory spatial data analysis with local statistics. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *47*(3), 415-421.

[2] Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, *27*(2), 93-115.

[3] Anselin, L. (2019). A local indicator of multivariate spatial association: extending Geary's C. *Geographical Analysis*, *51*(2), 133-150.

[4] Oxoli, D. (2019). *Exploratory approaches in spatial association analysis: methods, complements, and open GIS tools development*. Doctoral dissertation, Politecnico di Milano, Italy.

[5] Oxoli, D., Prestifilippo, G., Bertocchi, D., & Zurbaràn, M. A. (2017). Enabling spatial autocorrelation mapping in QGIS: The Hotspot Analysis Plugin. *GEAM. Geoingegneria Ambientale e Mineraria*, *151*(2), 45-50.

Analysis of OSM data through OSM-Notes user posting

Toshikazu Seto[1,*], Hiroshi Kanasugi[1] and Yuichiro Nishimura[2]

[1] Center for Spatial Information Science, The University of Tokyo, Tokyo, Japan;
tosseto@csis.u-tokyo.ac.jp, yok@csis.u-tokyo.ac.jp
[2] Division of Humanities and Social Sciences, Nara Women's University, Nara, Japan;
nissy_yu@cc.nara-wu.ac.jp

* 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.

Data from Volunteered Geographic Information (VGI) and public surveying are essentially different; while a great quality and quantity of data is available for VGI, discussions about data quality are scarce [1]. As the use of VGI in business expands, attempts have been made to develop quality verification tools (e.g. KeepRight and Validation OSM) with the understanding that, in addition to position accuracy, data quality equally depends upon the diversity and interaction of the number of users involved in data generation [2].

OSM-Notes is a capability for describing errors and discussing OSM data and fixme, although there are few research cases [3]. This feature is new and was only added to the openstreetmap.org website in April 2013. It allows users to specify and comment on any point on the map, and the history of comments can be accumulated and closed when the problem is solved. Unlike the fixme tag (allowing contributors to mark objects and places that need further attention in the form of a "note to self" or request for additional mapping resources), this function does not directly associate with OSM data, and there is no need to have an OSM account to create one. In this research, the OSM-Notes feature is mainly viewed as data that can be examined speedily from OSM data globally in terms of the content of the notes posted and the location of users. The purpose of the research is to analyze the context of OSM-Notes for quantitative and qualitative GIS approach [4].

This analysis was conducted on data dumped from Planet OSM as April 2019. Since planet-notes.osn (about 782 MB) is a special binary format, we used an enhanced parser able to separately output open (unresolved discussions) and closed (resolved discussions) based on "osn2osm". After converting to .osm format data, the set was combined with Natural Earth's border data and population data and treated as spatial data. As a result, there were OSM-Notes postings in 237 countries: 415,433 open and 129,887 closed records.

By counting the number of OSM-Notes postings by country based on whether they are open or closed, we determined that the majority were in the United States, Germany, and Russia. Moreover, it became clear that Japan, Canada, Korea, and Taiwan are the regions where there are a large number of submissions in urban areas (based on Natural Earth's

Seto et al. (2019). Analysis of OSM data through OSM-Notes user posting
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, 15-16. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387685

definition). In addition to these, many OSM users have posted from many countries, including Iraq, Ukraine, and Ecuador, and it is clear that active discussions are being held by many contributors. In these countries, mapping for humanitarian assistance is commonplace, as other reference resources are unavailable, so it is necessary to improve data quality through the use of OSM-Notes. OSM-Notes even has a function that allows non-OSM users to post, as is the case with 50% or more of the posts in Spain, Korea, etc.

This data can be used to analyze urban trends and spatial features within a single country. For example, according to the analysis for Japan, OSM-Notes has many posts about the location of shops and POIs (Points of Interest), and suggestions based on Maps.ME, Facebook, and Pokémon Go. This is considered to be the main reason that anonymous posting is permitted. It is also worth noting that very few users post only a single note. Thus, by analyzing OSM-Notes, it is possible to grasp hot issues between users of OSM data.

Overall, we found that postings to OSM-Notes are more frequent in cities where OSM mapping is active. On the other hand, how to accept bug reports from applications other than OSM (and not users' discussions) which edit areas where OSM data is limited remains an open problem. We need to consider the OSM data update process for multiple applications. Moreover, how to accept it is an important issue in considering the process of constructing OSM data. However, because the functional relationship between OSM-Notes and the data on OSM cannot be specified directly, the nature of the feature is also difficult to grasp directly compared to the fixme tag.

References

[1] Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, *31*(1), 139-167.

[2] Haklay, M. (2010). How good is volunteered geographical information? A comparative analysis of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design, 37*(4), 682-703.

[3] Seto, T., Iwasaki, N., & Nishimura, Y. (2017). Evaluation of the data update frequency and user interaction of OSM in Japan: case study of OSM-Note. In *State of the Map 2017*.

[4] Cope, M., & Elwood, S. (2009). *Qualitative GIS: A Mixed Methods Approach*. SAGE publications, London, UK.

Corporate Editors in the Evolving Landscape of OpenStreetMap: A Close Investigation of the Impact to the Map & Community

Jennings Anderson[1,*], Dipto Sarkar[2] and Leysia Palen[1,3]

[1] Department of Computer Science, University of Colorado Boulder, Boulder, CO, United States; jennings.anderson@colorado.edu
[2] Department of Geography, National University of Singapore, Singapore; dipto.sarkar@mail.mcgill.ca
[3] Department of Information Science, University of Colorado Boulder, Boulder, CO, United States; leysia.palen@colorado.edu

* 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.

OpenStreetMap (OSM) is both a map and the active community of over a million mappers that create and maintain it. Participation in OSM has largely been studied in terms of motivation and the resulting data quality. Today, the community is comprised by many different interest groups including craft/hobby mappers, humanitarian mappers, professional mappers, and more. The last few years have seen a dramatic growth in a specific group of mappers: corporate editors. These are mappers hired by corporations and edit the map as part of their employment. In November 2018, the OSM Foundation published the *organized editing guidelines* that outline a number of steps all groups engaged in organized editing activities (including corporate data teams) should take to promote transparency, openness, and engagement with other mappers—especially local community. This work identifies ten corporations that are complying with these guidelines and explores their mapping activities. We found these corporations have cumulatively edited over 17M objects globally in the last five years, of which 9M were edited in 2018 [1].

First, we traced the history of corporate involvement in OSM to show that while this growing phenomena of corporate editing is new, it represents just the latest stage in a long history of corporations both contributing to and benefitting from OpenStreetMap. Next, we used historical quarterly-snapshot OSM-QA-Tiles to quantify where the ten corporations are active on the map and what types of edits they are performing. We find these edits are global in geographical scope, yet vary per corporation in location and edit type: Corporations heavily impact road networks, yet non-corporate mappers maintain the majority of all edits by mapping more buildings and points-of-interest [1]. To date, this research has quantified and contextualized the growing phenomena of corporate editing in OSM and identified the need for more in-depth analysis to more descriptively explain the impact to the map and volunteer mappers in these regions where corporate-editors are active.

Anderson et al. (2019). Corporate Editors in the Evolving Landscape of OpenStreetMap: A Close Investigation of the Impact to the Map & Community
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, 17-18. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387693

To further explore these impacts, we need to dig deeper into the editing record to describe the evolution of the map. For this, we are building upon open-source OSM data-processing tools to construct new vector tiles with with full OSM editing histories [4]. These new historical analysis tiles allow us to efficiently explore the evolution of the map in these regions. This allows us to better contextualize and visualize the interactions between corporate editors and volunteer mappers at scale. Previous research has shown that the road network typically gets mapped first and the map builds up from there [2]. To this extent, we will explore the notion of *map seeding* whereby paid editors create the first version of the road network, seeding the map for others to maintain and grow. Supporting such an idea is the concept of *densification* of the map, where some mappers prefer to edit where there is existing—though incomplete or sparse—map data, instead of a beginning with a blank section of map [3]. The concept of such editing patterns highlights the nuances in effectively measuring the impact of paid editing on the map. In other words, this question is more complicated than "are corporate editors taking over?"

The first part of the research presented here is covered in [1]. The deeper exploration of the data to identify and explain the impact to the map and local communities requires first reconstructing the complete editing history of the map in the regions where corporate editors have been active. Specifically, this involves classifying and describing the interactions between two subsequent editors to expand the definitions of the terms *map seeding* and *densification* within the map. Given the global scale of corporate editing, the volume of OSM data, and the multitude of ways that contributors can edit the map, investigating these interactions between corporate and non-corporate contributors is a multi-faceted problem within Geoinformation Science.

References

[1] Anderson, J., Sarkar, D., & Palen, L. (2019). Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *8*(5), 232.

[2] Ciepłuch, B., Mooney, P., & Winstanley, A. C. (2011). Building Generic Quality Indicators for OpenStreetMap. In *Proceedings of the 19th Annual GIS Research UK (GISRUK)*. Retrieved from http://mural.maynoothuniversity.ie/2483/1/Gisruk2011-edit2-Blazej.pdf

[3] Corcoran, P., Mooney, P., & Bertolotto, M. (2013). Analysing the growth of OpenStreetMap networks. *Spatial Statistics*, *3*, 21-32.

[4] OSM-Wayback. Retrieved from at https://github.com/osmlab/osm-wayback

OpenStreetMap as Space

Dipto Sarkar[1,*] and So Hoi Kay[1]

[1] Department of Geography, National University of Singapore, Singapore; dipto.sarkar@mail.mcgill.ca, e0202937@u.nus.edu

* 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.

OpenStreetMap (OSM) is arguably the largest repository of crowdsourced geospatial information. However, much of the published literature regarding OSM has remained comfortably within the quantitative boundaries of data-driven inductive research. The interpretation of OSM as a product of contributions, created by a community of contributors, has governed the two major pathways of research. First, research concentrates on the current and changing nature of the OSM contributing community, with insights derived regarding the types of user profiles, motivations of contributors and the frequency of contributions. Second, studying OSM as a database, derives knowledge about the quality and quantity of data, ranging from the inequalities in representations to topological inaccuracies presented in the data.

The dominance of quantitative studies has led to a myopic and restrictive focus on OSM data and users with little discussion on the intertwining of the two. The aim of this research is to offer solutions to potentially address this gap by demonstrating how postmodern concepts from sub-disciplines in Human Geography are prevalent in the OSM landscape. Through this, we demonstrate how a postmodernist approach qualifies OSM as a space of interest to be studied in the field of Human Geography.

The evolution of the OSM landscape over time has demonstrated a manifestation of various postmodern geographical theories, concepts, perspectives and paradigms. Reflecting the transplantation of theories from well-established sub-disciplines of Human Geography, bears testament to OSM's worth as a space of interest to human geographers. The fluid, dynamic and ever-changing nature of OSM adds to the complexities of the landscape and qualifies it as a space to study new trends, emerging theories and re-negotiations of ideas. The aforementioned attributes of the OSM landscape render it suitable to be considered as a microcosm of geographical landscapes. Contrary to positivist approaches, Human Geography places emphasis on the representations, meanings, values, and intentions within spaces.

To demonstrate the credibility of OSM as a space, we exemplify how some concepts of Human Geography can be used in OSM:

1. Actor Network Theory (ANT): ANT calls for the understanding of the constant power negotiations and shifting networks of relationship between human and non-human actors. ANT is useful in studying the roles of different actors in the OSM landscape by drawing focus onto how things are "stitched together" across divisions and

Sarkar and Kay (2019). OpenStreetMap as Space
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, 19-20. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387699

distinctions, such as the differing intentions, profiles, and levels of authority among the OSM contributors [1].

2. Foucault's theory of discourse and power: In this theory, knowledge is perceived to be embedded within power networks. In OSM, discourse refers to the mapping edits, while the reality is reflected in OSM map. Edits vary according to the intersectionality and power of each user, which shapes OSM data. This should be taken into consideration in the analysis of OSM.
3. Urban Geography: Within urban geography, the postmodernist approach of participatory planning, that emphasizes inclusivity, collaboration, representation and a decentralisation. In the OSM context, it is the right to participation or awarding power to individuals, that needs highlighting, especially considering the barriers faced by certain marginalized groups. Several collaborative approaches in OSM has sought to narrow this gap.
4. Neoliberalization: 'Accumulation by Disposition' elucidates the phenomenon whereby originally public spaces, also professed as democratic and free spaces for public use, are deliberately converted to private spaces through the gradual dominance and uptake of spaces by private entities. In a manner, this phenomenon can potentially unfold in the OSM landscape, whereby a formerly 'public space', or in this case, a 'free-to-access digital space' is experiencing a spike in neoliberal intentions and presence with the increased corporate interest in OSM [2].
5. Gender Geographies: Gender geography brings to light the production and reproduction of inequalities of power and representation between men and women in various spaces [3]. In OSM, analysis has unveiled the similar production of inequalities, in terms of first, the proportion of genders which constitute the mapping community, and secondly, the resultant inequalities observed in the data. Utilization of concepts learnt in gender geography can mitigate issues of gender misrepresentation.

Our overarching aim is to explore the possibility of using OSM in classroom teaching as an instance of a geographic space. OSM is a particularly interesting case study as geography students are familiar with the geospatial data and its map based representation. Further, the crowdsourced nature in which the data is produced brings to the forefront interaction between entities, leading to complex dynamics. We provide some instances of how OSM as space may be of interest to geographers. Concepts of Human Geography manifested in OSM can be used to understand how digital geographies and offline activities are intrinsically interwoven. Thus, introduction to digital geography in classrooms can be through using OSM as a space to demonstrate how our understanding of offline world manifests in online spaces.

References

[1] Budhathoki, N. R., & Haythornthwaite, C. (2013). Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap. *American Behavioral Scientist*, *57*(5), 548-575.

[2] Anderson, J., Sarkar, D., & Palen, L. (2019). Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *8*(5), 232.

[2] Stephens, M. (2013). Gender and the GeoWeb: divisions in the production of user-generated cartographic information. *GeoJournal*, *78*(6), 981-996.

Assessing the Completeness of Urban Green Spaces in OpenStreetMap

Christina Ludwig[1,*], Robert Hecht[2], Sven Lautenbach[3], Martin Schorcht[2] and Alexander Zipf[1]

[1] GIScience Research Group, Heidelberg University, Heidelberg, Germany; christina.ludwig@uni-heidelberg.de, zipf@uni-heidelberg.de
[2] Leibniz Institute of Ecological Urban and Regional Development, Dresden, Germany; r.hecht@ioer.de, m.schorcht@ioer.de
[3] HeiGIT (Heidelberg Institute for Geoinformation Technology), Heidelberg University, Heidelberg, Germany; sven.lautenbach@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.

Urban green spaces provide a variety of important ecosystem services such as micro-climate regulation, increase of biodiversity and the provision of recreational and cultural services for citizens. Thus, they are an important factor for the quality of life in cities [1]. However, in order to take advantage of these services citizens need to have sufficient information about the location and qualities of urban green spaces. Within the project “meinGrün” (funded by the German Federal Ministry of Transport and Digital Infrastructure, BMVI) we are addressing this issue by developing a web-based recommendation service which helps citizens find nearby green spaces that satisfy their personal needs.

OpenStreetMap (OSM) plays an important role in this project, since it provides a lot of valuable information about urban green spaces and the amenities they provide (playgrounds, benches, etc.). However, the spatially heterogeneous data quality of OSM, especially in regard to the level of completeness, provides challenges for its usage in a recommendation system. Therefore, the integration of OSM data for our purposes requires a prior assessment of the completeness of urban green spaces. The level of completeness of certain geographic objects is one of the main fields of investigation in regard to OSM data quality. In recent years several studies investigated the completeness of OSM data with respect to the road network [2], buildings [3] or land use features [4]. Urban green spaces, on the other hand, were rarely the focus of completeness studies. Ali et al. [5] developed a method to quantify the plausibility of vegetation-related tags being assigned to specific OSM features and Lopes et al. [6] evaluated the potential of OSM for extracting information about natural local climate zones. Since both of these studies do not explicitly address the completeness of urban green spaces, we developed a new methodology for this purpose.

In contrast to buildings and highways this poses unique challenges due to the variety of vegetation-related OSM tags and the many different forms of urban vegetation ranging from large parks over private gardens to roadside greenery. OSM tags that describe natural objects are numerous and sometimes conceptually overlapping e.g. some features could be tagged as *leisure=park* or *leisure=garden*. This leads to different representations of urban

Ludwig et al. (2019). Assessing the Completeness of Urban Green Spaces 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, 21-22. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387701

green spaces in OSM across different geographical regions. Defining just one set of relevant OSM tags to measure the completeness of urban green spaces that can be applied everywhere is therefore not possible. Furthermore, the high degree of uncertainty that comes with this vagueness needs to be taken into account when using the data.

We developed a method to intrinsically measure the completeness of public urban green spaces based on the Dempster-Shafer Theory of Evidence (DST). First, a joint analysis of OSM data and the Normalized Difference Vegetation Index (NDVI) derived from Sentinel-2 imagery was used to quantify the relationship between a certain OSM tag and probability for the presence of vegetation. Urban green spaces marked with tags that were associated with high NDVI values were considered as explicitly mapped in OSM. In order to quantify the completeness of these data, additional indicators for the presence of public urban green spaces were derived based on the following geographic context variables: the density of foot paths, the presence of a POI (playground) and high NDVI values. These indicators were combined using DST to get a second estimate for the presence of public urban green spaces. Comparing this evidence to the map of explicitly tagged public green spaces in OSM yielded an intrinsic measure for completeness along with an estimate of its uncertainty. As a basis for this comparison the study area was divided into patches of homogenous land use based on natural and human-made barriers such as the road network, rivers or objects that mark changes in land use (fences, walls, etc.).

Results for the City of Dresden show that most public green spaces (e.g. municipal parks) are mapped with a high degree of completeness. However, publicly accessible, but privately owned green spaces within residential areas (e.g. courtyard of an apartment building with a playground) are only sometimes explicitly mapped as green spaces using tags such as *landuse=grass*. By comparing explicitly tagged public green spaces to the ones only indicated by geographic context (high density of foot paths, presence of a playground and high NDVI values) it was possible to successfully identify building blocks that contain public green spaces not explicitly mapped in OSM. In addition, these results were also compared to official data from the City of Dresden.

References

[1] Bolund, P., & Hunhammar, S. (1999). Ecosystem services in urban areas. *Ecological economics*, *29*(2), 293-301.

[2] Barrington-Leigh, C., & Millard-Ball, A. (2017). The world's user-generated road map is more than 80% complete. *PloS one*, *12*(8), e0180698.

[3] Hecht, R., Kunze, C., & Hahmann, S. (2013). Measuring completeness of building footprints in OpenStreetMap over space and time. *ISPRS International Journal of Geo-Information*, *2*(4), 1066-1091.

[4] Jokar Arsanjani, J., Mooney, P., Zipf, A., & Schauss, A. (2015). Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets. In *OpenStreetMap in GIScience*, 37-58. Cham: Springer.

[5] Ali, A., Sirilertworakul, N., Zipf, A., & Mobasheri, A. (2016). Guided classification system for conceptual overlapping classes in OpenStreetMap. *ISPRS International Journal of Geo-Information*, *5*(6), 87.

[6] Lopes, P., Fonte, C., See, L., & Bechtel, B. (2017). Using OpenStreetMap data to assist in the creation of LCZ maps. In: *2017 Joint Urban Remote Sensing Event (JURSE)*, 1-4.

Client-side route planning: preprocessing the OpenStreetMap road network for Routable Tiles

Harm Delva[1,*], Julián Andrés Rojas Meléndez[1], Ben Abelshausen[2], Pieter Colpaert[1] and Ruben Verborgh[1]

[1] IDLab, Ghent University – imec, Ghent, Belgium; harm.delva@ugent.be, julianandres.rojasmelendez@ugent.be, pieter.colpaert@ugent.be, ruben.verborgh@ugent.be
[2] Open Knowledge Belgium, Brussels, Belgium; ben@openknowledge.be

* 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.

Travelers have higher expectations than current route planning providers can fulfill, yet new solutions struggle to break through. Matching user experience from existing applications is already challenging without the large-scale infrastructure most of them have at their disposal; additionally integrating datasets such as the road network, public transportation schedules, or even real time air quality data is an even more laborious endeavour.

The World Wide Web Consortium (W3C) and the Open Geospatial Consortium (OGC) mention the usage of Linked Open Data as a best practice for publishing interoperable geospatial datasets. Instead of relying on proprietary data formats or monolithic CSV files, Linked Open Data uses the Resource Description Framework (RDF) data model as a framework for existing domain models. Every data element, and even the relations between them, receives a Uniform Resource Identifier (URI). Data publishers can reuse these identifiers to unambiguously refer to resources on the Web, thus making individual data sets more interoperable. The ultimate goal being automated integration, giving even clients the power to execute the queries. Client-side querying differs from traditional approaches but provides some advantages: (i) it takes the load off the service provider, (ii) the data can be cached for subsequent queries, (iii) the user leaks less personal data.

The OpenStreetMap road network has recently been published as routable Linked Open Data, following a similar approach to vector tiles [1]. However, executing route planning queries on the client is still an unsolved problem. Long-distance queries require large amounts of data and downloading all the data takes a long time. State-of-the-art route planning algorithms achieve better query execution times by using auxiliary data that has been computed in a preprocessing phase [2, 3]. The biggest bottleneck in client-side querying is the network; downloading more data to improve query times will ultimately make querying even slower. Client-side route planning requires a different approach to match the quality of service of existing services.

Delva et al. (2019). Client-side route planning: preprocessing the OpenStreetMap road network for Routable Tiles
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, 23-24. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387705

We explore several ways of preprocessing the routable tile data to improve the *user-perceived performance* of query evaluation. Each tile can be processed independently from the others, resulting in a pleasingly parallel workload. As a first step, we compute how to efficiently traverse pedestrian areas. Only the boundary edges of these areas are defined in OpenStreetMap which means that most routing engines route along these edges, yielding suboptimal paths. This step materializes the information that's already present in the base data to make it easier for route planners to ingest.

Secondly, we identify which nodes and ways are actually needed to cross a specific tile, filtering out the elements that are only used for local traffic. Queries only need the full tiles around the departure and arrival locations. This process becomes more effective at higher zoom levels: at the lowest zoom level it decreases the file sizes by roughly 50% and this increases to 86% at higher zoom levels. Finally, we discard all untagged nodes that aren't used to connect ways – replacing them with precomputed distances between the remaining nodes. This reduces the file size by another 50%, regardless of the zoom level.

Each step yields a different view of the tile data and the results are published as Linked Open Data, in accordance with the W3C and OGC best practices. Similar preprocessing steps as the ones described in this paper are often done while setting up a route planning service, with a significant amount of duplicate work as a result. This process could be greatly simplified if the resulting data is freely accessible.

We integrated the resulting datasets into a route planner for public transportation that uses Dijkstra's algorithm for the road network pathfinding. All data are downloaded dynamically as they are needed. We found that short-distance one-to-many queries such as finding the closest nearby bus stops that initially took around 400 ms to complete only take around 260 ms with the preprocessed data, and the first results are presented after 140 ms. The difference becomes bigger over long distances; computing a 50 km and a 100 km journey used to take respectively 7 minutes and 56 minutes. Using the preprocessed data lowers the query times to respectively 10 s and 21 s. We see similar improvements in the amount of downloaded data; these journeys initially required 30 MB and 170 MB of data, which is reduced to 3 MB and 9 MB.

Our contributions in this paper are two-fold; we have developed preprocessing techniques driven by our need to make a serverless route planner more efficient and in doing so discovered that the challenges we faced are not unique – and that we as a community could benefit from sharing more data with each other.

References

[1] Colpaert, P., Abelshausen, B., Rojas Meléndez, J., Delva, H., & Verborgh, R. (2019). Republishing OpenStreetMap's roads as Linked Routable Tiles. In *Proceedings of the 16th ESWC: Posters and Demos*.

[2] Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, *18*(6), 877-895.

[3] Geisberger, R., Sanders, P., Schultes, D., & Delling, D. (2008). Contraction Hierarchies: Faster and Simpler Hierarchical Routing in Road Networks. In *Proceedings of the Seventh International Conference on Experimental and Efficient Algorithms*, 319-333.

Estimating latent energy demand of buildings

Nikola Milojevic-Dupont[1,2,*], Peter-Paul Pichler[3], Lynn H. Kaack[4], Steffen Lohrey[2] and Felix Creutzig[1,2]

[1] Land use, Infrastructure and Transport, Mercator Research Institute on Global Commons and Climate Change (MCC), Berlin, Germany; milojevic@mcc-berlin.net, creutzig@mcc-berlin.net
[2] Sustainability Economics of Human Settlements, Technische Universität Berlin, Berlin, Germany; steffen.lohrey@campus.tu-berlin.de
[3] Social Metabolism & Impacts, Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany; pichler@pik-potsdam.de
[4] Energy Politics Group, ETH Zürich, Zürich, Switzerland; lynn.kaack@gess.ethz.ch

* 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.

Buildings are responsible for a quarter of energy-related greenhouse gas emissions, making them a substantial driver of climate change; heating and cooling in buildings in particular are the biggest contributors to their energy demand [1]. Strategies that reduce energy demand in the building sector also need to be socially just by ensuring that minimal thermal comfort is accessible to all. Of particular concern is the energy use of cooling, as deadly heat waves are expected to become more frequent in many regions of the world, making cooling necessary to ensure basic wellbeing [2].

In order to develop mitigation solutions at scale, a key challenge is to understand the energy demand across the large and heterogeneous building stock. However, building energy data are often proprietary or do not even exist – which hinders large scale studies. Current methods are well-equipped to model individual buildings but do not scale up easily, or they model the overall building sector – leaving out much of the local context. This confines impactful climate action to a limited group of cities, where data exists. There are many regions with pressing climate change mitigation and development challenges that are often overlooked, in particular rapidly-urbanizing urban areas in the Global South [3].

Our project aims at leveraging the increasing availability of spatial data, such as OpenStreetMap (OSM) data, and machine learning techniques, to estimate the minimum – latent – energy demand for heating and cooling in a larger building stock. Our framework is data-driven and modular so that it can be made more complex as more data become available. We use first-order factors influencing latent energy demand, which are more easily available at scale.

The outer wall surface is particularly important and requires a 3D model of the building stock. The coverage of 2D building footprints in OSM is rapidly increasing. Even though some issues remain, in many cities the data quality is likely to be sufficient for our purpose. However, the height information is sparsely populated in OSM data, and with large imbalances between regions: only ~10 million buildings worldwide have values for the height key in July 2019. This corresponds to about 20 cities of the size of Berlin fully mapped. Best

Milojevic-Dupont et al. (2019). Estimating latent energy demand of buildings
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, 25-26. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387711

available height data is provided either by cadaster data, or by an expensive aerial sensing technique – LiDAR – which is only openly available for very few cities in developed countries.

Our overall workflow is divided into two main models: (i) a building height prediction model and (ii) a model of the energy demand for heating and cooling individual buildings to a given temperature.

Here, we present the concept for the building height prediction model. We predict buildings heights using supervised machine learning techniques that map the relationships between buildings and urban tissue features from OSM to building heights from ground truth LiDAR data. While a similar approach has been taken for two Dutch cities by [4], here we aim at developing a model that can predict buildings heights from OSM data only in order to scale to a larger building stock.

Preliminary results show that OSM data alone are predictive of building heights. We are currently developing our model on a set of European cities where open 3D data are available. From preliminary results for Berlin data, we find that heights can be predicted with a mean average error of ~3 meters using a random forest regression with 5-fold cross-validation. To improve the prediction, we are investigating deeper architectures and further inclusion of the surroundings of the building under scrutiny. This includes engineered features like street connectivity, and rasterized OSM building and street layers.

One of the main challenges of this model is to generalize to new cities, especially in different geographical regions. It can be expected that the relationship of certain building features with height is fairly similar across the world, while that of others differ widely. Transfer learning with local retraining may enable to take advantage of those regularities, while adapting to local specificities.

The building height attribute in OSM data provides crucial information for climate solution research in cities. Machine learning can partly infer missing data, but algorithms need training data to learn from. OSM mappers can help provide this data, or other relevant predictors like the number of floors in a building. Information about only a limited amount of buildings could be enough, if those data points are well distributed across cities and regions. We further emphasize the general need for detailed infrastructure data: by better supporting municipal policy makers, data-driven urban planning strategies have a great potential to mitigate climate change in cities.

References

[1] Intergovernmental Panel on Climate Change (2018). Global warming of 1.5°C. An IPCC special report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Retrieved from https://report.ipcc.ch/sr15/pdf/sr15_spm_final.pdf

[2] Isaac, M., & Van Vuuren, D. P. (2009). Modeling global residential sector energy demand for heating and air conditioning in the context of climate change. *Energy policy*, *37*(2), 507-521.

[3] Creutzig, F., Agoston, P., Minx, J. C., Canadell, J. G., Andrew, R. M., Quéré, C. L., Peters, G. P., Sharifi, A., Yamagata, Y., & Dhakal, S. (2016). Urban infrastructure choices structure climate solutions. *Nature Climate Change*, *6*(12), 1054-1056.

[4] Biljecki, F., Ledoux, H., & Stoter, J. (2017). Generating 3D city models without elevation data. *Computers, Environment and Urban Systems*, *64*, 1-18.

Towards Scalable Geospatial Remote Sensing for Efficient OSM Labeling

Rui Zhang[1], Marcus Freitag[1], Conrad Albrecht[1,*], Wei Zhang[1] and Siyuan Lu[1]

[1] Data Intensive Physical Analytics, TJ Watson Research Center, IBM Research, Yorktown Heights, NY, United States; rui.zhang@ibm.com, mfreitag@us.ibm.com, cmalbrec@us.ibm.com, weiz@us.ibm.com, lus@us.ibm.com

* 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.

The past decade has shown a dramatic increase in the amount of openly available geo-spatial datasets such as multi-spectral and RADAR satellite imagery from space agencies like ESA and NASA government-sponsored, high-resolution aerial survey raster data, e.g., USDA (U.S. Department of Agriculture) weather reanalysis model data based on sensor networks, e.g. published by the PRISM Climate group geo-tagged messages as well as images from social media platforms such as Twitter and Instagram, etc. accumulating geo-tagged information at data rates easily exceeding tens of terabytes a day.

Given that an open database project such as OpenStreetMap (OSM) relies on volunteers to spend their valuable time to generate vector datasets that annotate and update information on roads, buildings, land cover, points of interest, etc., it is natural to ask how sources of freely available spatio-temporal information might help to support and guide mappers in their work.

At the same time, major progress has been made in the open-source digital arena of big data processing and artificial intelligence (AI). For example, projects for distributed non-relational database systems such as HBase (https://hbase.apache.org/) or in-memory distributed compute frameworks such as Spark are available to run on commodity hardware to scale analytics. Deep learning libraries such as PyTorch (https://pytorch.org) in accordance with the explosive amount of neural network architectures published by academia enable, for example, state-of-the-art computer vision algorithms which can be leveraged for remote sensing tasks: detection of buildings, land classification, change detection, etc.

Our work discusses and demonstrates how to link tools from big data analytics and machine learning to geo-spatial datasets at scale in order to extract value from openly available spatio-temporal datasets to the potential benefit of OSM mappers. In particular, we show the design of a system that employs the key-value store HBase to index spatio-temporal satellite imagery to let Spark-SQL (https://spark.apache.org/sql/) user-defined functions act on it to remotely identify human signatures on Earth's surface by the aid of AI.

Zhang et al. (2019). Towards Scalable Geospatial Remote Sensing for Efficient OSM Labeling
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, 27-28. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387715

Finally, when it comes to pixel-wise land classification, we are using the 1-meter resolution USDA aerial survey data and information derived from the OpenStreetMap project. The goal is to establish a scalable pixel level translation model from aerial map to OSM, where colors and shapes define land classification, i.e., forestry, grassland, building, road, etc. The USDA aerial survey is refreshed every other year, so we expect to translate the latest aerial survey to OSM and compare with the current OSM state to identify changes on the actual land use. This information will guide the OSM community where the map needs to be updated.

We believe that the techniques and use cases presented will help to identify hot spots of where OSM needs human labor most – either in mapping or updating labels. Moreover, we hope to spark a scientific, strategic and technical conversation with the OSM community on needs regarding semi-automated support systems for global mapping. As a bonus, we will introduce the open-source tool [1] to interact with the spatio-temporal platform PAIRS that supports our research.

References

[1] IBM (2019). IBM PAIRS Geoscope. Retrieved from https://github.com/IBM/ibmpairs

Workforce Development and YouthMappers: Understanding perceptions of students in humanitarian mapping

Patricia Solís[1,*] and Sushil Rajagopalan[2]

[1] YouthMappers, Knowledge Exchange for Resilience, Arizona State University, Tempe, AZ, United States; patricia.solis@asu.edu
[2] School of Sustainability, Arizona State University, Tempe, AZ, United States; srajag20@asu.edu

* 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.

Over the last few decades, the geospatial workforce has grown globally across all sectors and the sector is expected to grow at a compound annual growth rate (CAGR) of 19% from 2017 to 2023 [1]. Humanitarian mapping efforts are one way for young people to gain valuable career preparation experience in both technical and practical aspects of the geospatial industry to take advantage of such opportunities. YouthMappers is a global network of local chapters, student-led groups based out of currently 157 university campuses in 42 countries around the world, which organizes open mapping for humanitarian and development needs while serving as a platform for capacity building [2].

While students prepare themselves for careers through tertiary education, their association with extracurricular activities such as YouthMappers may help them acquire certain geospatial skill sets as well as other "soft" job skills that are critical for this growing job market because of its connection to authentic, real-world applications, and a framework that is explicitly linked to geospatial competencies and learning objectives [3, 4]. From preliminary assessments and anecdotal experience, we know that the humanitarian purpose of the data creation efforts motivates YouthMappers students to participate, but does it lead to better preparation for a global workforce? Which activities and engagement mechanisms correspond to perceptions of being better prepared? How do these activities differ by gender or location? What do mappers gain from their volunteering? Does it matter if their engagement includes formal coursework at all or in a particular field or might it be sufficient to engage in informal types of activities of YouthMappers alone? We will present an analysis of results from an online survey of YouthMappers around the world to understand the perceptions on career preparedness of the students who have been engaged in humanitarian mapping with YouthMappers.

Our study indicates that, on average, respondents report having experience with using 3.39 different geospatial tools, where males have a mean of 3.45 and females, 3.26. One-way ANOVA tests shows that there is a significant difference in the mean number of

Solis and Rajagopalan (2019). Workforce Development and YouthMappers: Understanding perceptions of students in humanitarian mapping
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, 29-30. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387717

geospatial tools used by YouthMappers [F (2,215) = 9.821, p = 0.000] among groups spending shorter versus longer time periods participating in YouthMappers activities. The percentage of students who consider their skill level as 'expert' in geospatial tools increases with the time spent participating in YouthMappers activities. Overall, 38 percent of students with 2 years or more participation in YouthMappers consider themselves as 'expert', while only 24.5 percent with 1 to 2 years do so; meanwhile 15.5 percent of those with less than 1 year participation opine the same. Almost 47 percent of YouthMappers feel "very prepared" for 'their professional career after they finish their degree,' and we found no statistically significant difference between male and female respondents. Results show that 65.7 percent of students "strongly agree" that 'their frequent participation with geospatial/mapping technology through YouthMappers make them a stronger candidate for employment'. Likewise, 62.3 percent feel that their YouthMappers experience has been "very helpful" in preparing students for a professional career, where answers were statistically significantly higher for female respondents at $p < 0.05$ level. The gender analysis reveals that 52 percent of females attribute being able to attend or participate in a national or international level conference, workshop, or meeting as a direct result of their participation in YouthMappers, compared to 38.7 percent of males. More than one-third of students indicate receiving an internship and/or job offer as a direct result of their YouthMappers experience.

The findings presented here confirm that the YouthMappers design contributes to key capacity building elements for students to prepare for geospatial careers, which include positive results for female mappers. The study indicates areas for further research and potential awareness-raising among participants about the value of extracurricular humanitarian mapping. By increasing geospatial skills among university/college students to prepare them for employment and careers, along with its efforts to eliminate gender disparities in acquiring such skills and deploying such activity in service to humanitarian and development purposes, YouthMappers makes important contributions to Sustainable Development Goal 4 (Quality Education) in a global context. On the basis of our findings, this paper hopes to stimulate discussions around the need to introduce such purposefully-designed extracurricular activities like YouthMappers in universities/colleges to both enhance the learning experience, and develop additional job market skills, in order to better prepare university/college students for working as global citizens in a geospatial workforce.

References

[1] Market Research Future (2019). Geospatial Analytics Market 2019. Retrieved from https://www.reuters.com/brandfeatures/venture-capital/article?id=101143

[2] Solís, P., McCusker, B., Menkiti, N., Cowan, N., & Blevins, C. (2018). Engaging global youth in participatory spatial data creation for the UN sustainable development goals: The case of open mapping for malaria prevention. *Applied Geography, 98*, 143-155.

[3] Hite, R., Solís, P., Wargo, L., & Larsen, T. (2018). Exploring Affective Dimensions of Authentic Geographic Education Using a Qualitative Document Analysis of Students' YouthMappers Blogs. *Education Sciences*, *8*, 173.

[4] Solís, P., Huynh, N. T., Carpenter, D., Adames de Newbill, M., & Ojeda, L. (2017). Using an authentic project based learning framework to support integrated geography education linked to standards and geospatial competencies. *Research in Geographic Education*, *19*(20), 36-65.

Automatic feature extraction to support Mountains Mapping in OSM

Rocio Nahime Torres[1,*], Piero Fraternali[1], Federico Milani[1] and Darian Frajberg[1]

[1] Department of Electronics Information and Bioengineering, Politecnico di Milano, Milan, Italy; rocionahime.torres@polimi.it, piero.fraternali@polimi.it, federico.milani@polimi.it, darian.frajberg@polimi.it

* 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.

Open access geographical databases, such as OpenStreetMap (OSM), offer a valuable alternative to proprietary solutions for the development of voluntary environment monitoring systems. However, the quantity and quality of information stored in such systems must be carefully evaluated and the contributions of volunteers must be boosted by means of effective engagement methods. We propose a hybrid approach, in which an open Digital Elevation Model (DEM) data set is processed with different techniques to find candidate mountain information and uncertainty in the automatically extracted candidates is reduced by means of voluntary crowd-sourcing. The improvement of landform information (not only about mountains, but also about orography and hydrography in general) can support the development of environment monitoring applications.

Different methods have been proposed to improve the generation of new objects in VGI, which usually rely on “in itinere” evaluation, hereby volunteers are given feedback while they are producing and annotating data. As an example, the work presented in [1] aims to improve the homogeneity of contributions by suggesting users tags while creating or editing objects, through a plug-in for Java OSM. Heterogeneity derives from the possibility that OSM provides to the users to create new tags at will. Consequently, OSM has now more than 101.000.000 different tags (https://taginfo.openstreetmap.org/reports/database_statistics), where more than one can refer to objects of the same class. As another example, the work presented in [2] aims at improving user’s contributions by means of rule-guided procedures that help volunteers classify grass-related features correctly. The authors developed the Grass&Green application, which warns the users about potentially incorrect classifications at editing time. The rules employed were derived with data-mining techniques. These approaches have an important advantage: evaluations are provided as feedback to the users, who can use them to confirm or discard their contribution; this enables immediate quality control, instead of deferred quality checking after the users have left the application. Nonetheless, the above mentioned approaches rely on the users creating their contributions from scratch and do not provide hints about the gaps in the existing information that could be a target for contributors.

Nahime Torres et al. (2019). Automatic feature extraction to support Mountains Mapping in OSM
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, 31-32. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387719

To tackle this, we propose a contribution framework that consists of three main steps: (1) automatic identification of candidate objects, (2) integration of the candidate objects into a validator for the crowd to confirm or discard objects, and (3) final inspection of the results to import them into OSM.

In our case, the objects of interest are mountain peaks. In OSM, such objects require the following mandatory information: id, latitude, longitude, version and timestamp, and can optionally host additional information such as name, elevation, and alternative names. There are different methods in the literature that can support peaks extraction from DEM data, which are mainly based on heuristics algorithms that encode the "definition" of mountain. In previous works, we also investigated, with success, the feasibility of using OSM data to train a Deep Learning algorithm that learns what is a mountain from the data [3].

To populate the crowd-sourcing tool with data, objects that the algorithms detected as candidate mountain peaks, but are not present in OSM, will be published for the crowd to validate. To reduce the number of false positives and the crowd work, only candidates found by multiple detectors (a Deep Learning method and a heuristic method present in Landserf) will be published. To decide if the two methods agree on a certain object, a distance threshold of 80 meters is employed. For example, in an area in the middle of the Swiss Alps (46°N-47N° and 8°E-9°E), where OSM presents 1493 mountain peaks, the two chosen detection methods found 656 mountain peaks missing from OSM. Another example could be in the area of Albania (40°N-41N° and 20°E-21°E), where OSM presents a very high number of mountain peaks (3952), and still the two methods find 156 mountain peaks missing from OSM. These two examples show the potential of our approach to find data missing in OSM.

The proposed approach aims at improving the completeness of OSM mountain peaks information by proposing to the crowd new mountain peaks found by already validated methods. The target audience could be passionate mountain enthusiasts, such as members of worldwide mountain clubs (e.g. the CAI: Italian Alpine Club) and the 400.000 active users of PeakLens, an AR mobile application for trekkers who are the perfect candidates to join a data collection campaign at the world scale.

References

[1] Vandecasteele, A., & Devillers, R. (2013). Improving volunteered geographic data quality using semantic similarity measurements. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-2/W1*, 143-148.

[2] Ali, A. L., Falomir, Z., Schmid, F., & Freksa, C. (2017). Rule-guided human classification of Volunteered Geographic Information. *ISPRS Journal of Photogrammetry and Remote Sensing 127*, 3-15.

[3] Torres, R. N., Fraternali, P., Milani, F., & Frajberg, D. (2018). A Deep Learning model for identifying mountain summits in Digital Elevation Model data. In *2018 IEEE First International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 212-217.

Characterizing player types in gamified geodata acquisition - An exploratory analysis of StreetComplete

Heinrich Lorei[1,*], René Westerholt[2] and Alexander Zipf[3]

[1] Metropolregion Rhein-Neckar, Mannheim, Germany; heinrich.lorei@m-r-n.com
[2] Centre for Interdisciplinary Methodologies, University of Warwick; rene.westerholt@warwick.ac.uk
[3] GIScience Research Group, Heidelberg University, Heidelberg, Germany; 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 Geographic Information (VGI) projects like OpenStreetMap (OSM) have recently gained wide interest also beyond specialist communities. Studies have shown, however, that the average OSM contributor is middle class or higher, typically middle age, and to a certain degree educated [1]. The concept of gamification is one way to widen and diversify user engagement in VGI projects. It allows making otherwise unattractive activities more compelling by using game design elements. The work presented addresses the use of gamification in the context of geodata collection.

The location-based Android application StreetComplete allows collecting attribute information for existing OSM features. A map view thereby indicates missing data in the vicinity of the players' current GPS positions, including names, speed limits, opening hours or roof shapes. Users collect this information on site by responding to questions, thus combining the less attractive task of data acquisition with playful geographic exploration. The main game design element used in this application is a scoring system that allows users to achieve points, and thus to compete with each other. It is the nexus between this scoring system with spatial and temporal parameters of gaming behaviours this work focuses on.

The main research question of this work is based on two assumptions. We assume that users are at least subconsciously aware that they are participating in a playful, competitive data collection; and we assume that certain parameters of users' gaming behaviour partly reflect their intention to succeed in the game. The latter is based on the first law of geography [2], and it implies assuming players to implicitly or explicitly employ "spatial strategies" based on scores obtained so far to determine which quest they shall solve next. Based on these assumptions, we address the following research question: To what extent is it possible to distinguish player types from the spatiotemporal parameters of their gaming behaviours?

The study presented uses a modified version of StreetComplete. A total of 40 test persons took part voluntarily, 28 of which were geography students at Heidelberg University.

Lorei et al. (2019). Characterizing player types in gamified geodata acquisition - An exploratory analysis of StreetComplete
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, 33-34. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387723

All test persons started the game from the same point and could independently explore a playing field located in Heidelberg. Location and extent of the study area were chosen to reflect realistic conditions with respect to OSM task variety and density, as real quests from the game were used. All GPS tracks were recorded, and then used to derive spatial and temporal parameters, such as game duration, distance travelled and average movement speed. The area of the standard deviation ellipse of each user's GPS positions further proxies the extent of their area covered, whereas the elongations of these ellipses are used as proxies for the target orientations of the users' movements. Another parameter calculated is a detour factor, given as the ratio between the actual and the "ideal" path length, the latter being the shortest path calculated using the OpenRouteService API.

Using cluster and archetype analysis, we can identify two general sorts of players: one group that shows interest in good performance concerning the scoring system, and a second group of players that seems more attracted by exploring the playing field. These results are in line with the widely regarded player classification scheme proposed for non-spatial settings where the two groups identified have been named "Achievers" and "Explorers" [3]. In addition, we were able to identify three novel player types: players who optimise for time ("grasshoppers"), players maximizing their number of tasks solved, regardless of point scoring ("grazers"), and those trying to integrate data collection efficiently with other trips ("en passant collectors"). These novel types of players are spatial and temporal in nature, indicating that geographic space is an important factor when it comes to characterising gamification, likely also beyond OSM. We thus expect the exploratory results presented to be of broad interest also to psychologists, cognitive, and social scientists, especially in view of the recent increase in interest these disciplines have shown in the use of geographical information and location-based techniques [4]. The identified player types can further be used practically to personalize StreetComplete and to advance the integration of further game elements.

References

[1] Jokar Arsanjani, J., & Bakillah, M. (2015). Understanding the potential relationship between the socio-economic variables and contributions to OpenStreetMap. *International Journal of Digital Earth*, *8*(11), 861-876.

[2] Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234-240.

[3] Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD Research*, *1*(1)

[4] Bluemke, M., Resch, B., Lechner, C., Westerholt, R., & Kolb, J. P. (2017). Integrating Geographic Information into Survey Research: Current Applications, Challenges and Future Avenues. *Survey Research Methods*, *11*(3), 307-327.

“Ohsome” OpenStreetMap Data Evaluation: Fitness of Field Papers for Participatory Mapping

Carolin Klonner[1,*], Maximilian Hartmann[1], Lily Djami[1] and Alexander Zipf[1]

[1] GIScience Research Group, Heidelberg University, Heidelberg, Germany;
c.klonner@uni-heidelberg.de, max_hartmann@stud.uni-heidelberg.de, lily@stud.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.

Crowdsourced information can support disaster management in different ways. Up-to-date information is provided by local citizens, which can be used to enhance flood models [1] or to capture flood risk perception [2]. Anyone with internet access can contribute new data to the collaborative mapping project OpenStreetMap (OSM), edit, and freely use the data provided by OSM. For an adequate use, the quality of the data needs to be known beforehand, and thus, the quality assessment of OSM is an important research topic. If reference data are available, extrinsic quality evaluation is possible [3]. However, often there is a lack of such resources. Therefore, intrinsic quality measures are required. Raifer et al. [4] developed the OpenStreetMap History Database (OSHDB) to analyze the temporal evolution and spatial heterogeneity of the OSM data at large scale.

Our study uses the OSHDB for analyzing OSM in Brazil, where we investigate flooding in São Paulo and Rio Branco within the Waterproofing Data Project. We apply the method of participatory mapping, which is based on OSM Field Papers [2]. Participants can mark their experiences on the Field Papers and these resulting Sketch Maps can automatically be georeferenced and thus, the collected data can be processed in a fast way. However, due to heterogeneous OSM data in different areas, it is necessary to evaluate the fitness for purpose of the Field Papers first. If the required base data are not available, it is difficult to use the Field Papers for the Sketch Maps. Participants might find it difficult, for example, to orientate themselves on the map due to a lack of the representation of important points of interest (POIs) and landmarks in OSM. Moreover, OSM data, which have not been updated for a long time, might be wrong and therefore can lead to errors in the drawing of the participants.

We developed a tool which bundles several intrinsic analyses to evaluate a study region's fitness for usage in Field Papers. Results are delivered in written form combined with recommendations for the person who wants to apply the Field Papers for participatory mapping. These recommendations include details to check beforehand and possible problems to be aware of during the usage of the Field Papers. The results are summed up by

Klonner et al(2019). “Ohsome” OpenStreetMap Data Evaluation: Fitness of Field Papers for Participatory Mapping
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, 35-36. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387725

a general score, which is displayed via a traffic light, to provide an easily accessible insight into the fitness of the OSM data. The analyses can be directly performed on multiple bounding boxes. In contrast to many other approaches, this tool does not require a local database to work on, but accesses the OSHDB via the ohsome-API developed by the HeiGIT. This is both time and resource saving for the user, who can directly use the program. The ohsome-API allows us to get already accumulated and calculated results for some requests like the density of features with certain tags. For other analyses we can access all features and versions of features that existed during the last year (of available data) with all their tags and coordinates. On these raw data, we can perform further operations which we cannot request directly. We can look at different versions of the same features, for example, to calculate the number and extent of positional changes.

The results of the evaluation are shown on a web page and the users can decide based on the given recommendations whether they want to apply the Field Papers. For a tested region in Heidelberg our results indicate a good fitness for usage in Field Papers (rated as green on the traffic light), and only one recommendation is given: every street and way feature was changed 0.17 times on average, which we consider as increased based on our experience, and which might indicate that some ways or streets are still mapped inaccurately; we advise users to be aware of this. In contrast, our results for Rio Branco indicate that there might be problems for the usage as the data are overall rated as yellow and more recommendations are necessary, e.g., users should take into account that streets and ways are possibly mapped inaccurately (average number of changes: 0.2), and explore how well participants can orientate themselves on the Field Papers because there is a low POI and landmark density (combined: 5.25 features per km²).

In conclusion, the presented tool enables detailed analyses of OSM data to support the decision making process in cases of the application of OSM Field Papers for participatory mapping. It offers a sustainable approach because local authorities, for example, can apply these methods themselves without expert knowledge. The presented tool will be integrated in a web portal which will facilitate the participatory mapping activities and present the data in combination with authoritative data. We are further developing the tool to allow for new forms of application.

References

[1] Assumpção, T. H., Popescu, I., Jonoski, A., & Solomatine, D. P. (2018). Citizen observations contributing to flood modelling: opportunities and challenges. *Hydrology and Earth System Sciences*, *22*(2), 1473-1489.

[2] Klonner, C., Usón, T., Marx, S., Mocnik, F.-B., & Höfle, B. (2018). Capturing Flood Risk Perception via Sketch Maps. *ISPRS International Journal of Geo-Information*, *7*(9), 359.

[3] Haklay, M. (2010). How good is volunteered geographical information? A comparative analysis of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, *37*(4), 682-703.

[4] 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).

Contextualizing OpenStreetMap in Mapping Favelas in Brazil

Everton Bortolini[1,*] and Silvana Philippi Camboim[1]

[1] Department of Geomatics, Federal University of Parana, Curitiba, Parana, Brazil;
evertonbertanbortolini@gmail.com, silvanacamboim@gmail.com

* 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.

About 6% of Brazilian population lives in Favelas [1]. They are characterized by their informality and precariousness of their physical structures, and by the social vulnerability of their population, as well as by their dynamics since they expand or change rapidly [2]. These characteristics impact on the outdated or lack of geographic information while there are demands for them to support the development of these communities [2].

Research that systematized knowledge about the mapping of spaces that are precarious and socially vulnerable, such as favelas, began in the 1990s. At first, the mapping techniques used were commonly referred to as the Participatory Geographic Information System (PGIS) [3] and Public Participation Geographic Information System (PPGIS) [4, 5]. From the technological context of the world until the mid-2000s, PGIS and PPGIS used analogic tools or Geographic Information Systems [6]. Web 2.0 provided new tools that can be applied to mapping favelas, such as OpenStreetMap [6]. However, it is still necessary to explore these tools within the limitations presented by the reality of the favelas. Therefore, this research aims to describe the collaborative mapping processes using OpenStreetMap in the context of favelas.

The notes that we made in this text were obtained from a case study with residents of six favelas (with a total population of about 15,000 people) and volunteers of a social organization - TETO Brasil, working in emergency housing construction and other activities in vulnerable communities. These communities are located in the Curitiba Metropolitan Region (Brazil), with about 2.5 million inhabitants. Thus, we applied a questionnaire to these volunteers to describe their profile and its relationship with the context of geographic information. In the case of residents we use data obtained by the NGO itself in its census process. In addition, to complement the research, we conducted a participant observation with these groups of people during the mapping activities conducted in 2017 and 2018. We analyze these data in order to first describe the actors and actions that influence them in the mapping process. In the analysis we consider both humans and nonhumans as actors.

As for the actors involved in the mapping processes with OpenStreetMap in favelas, we highlight the role of mediators. They are intended to support citizens in the use of tools and devices for the execution of mapping processes. Citizens of favelas, agents of NGOs, universities or other organizations can act as mediators. However, they need a previous

Bortolini and Camboim (2019). Contextualizing OpenStreetMap in Mapping Favelas in Brazil
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, 37-38. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387727

training of the mapping process, regardless of whether an internal or external actor to the slums. In the case of mediators, even if they have mediation skills with communities, it is not common to have the same skills for technical mediation in cartography.

In the mapping of slums, besides the mediators, other actors are the citizens. Three notes can be made about their participation. The first is the motivation to start a mapping process. We realize that the motivating factor must be linked to an activity with potential benefit to the community, such as an infrastructure project. The second note is the representativeness of the group. It makes the map include a meaningful view of the community and is influenced by the first item described, motivation. In practice, it is seen that this participation is not obtained homogeneously in all favelas. The third note is that participation is conditioned to space and time in which the mapping occurs. In favelas, these may occur in community meeting places or open public spaces in these communities. In the existence of support of external actors, there is the possibility of mapping in a remote and sometimes asynchronous way.

In the case of favelas, smartphones are widely used by the population, being a viable and important tool for mapping. They are used by approximately 50% of residents to access the internet while other devices at most 20%. In addition, paper maps are frequently used tools in mappings. However, because of their lack of knowledge of these tools by citizens and mediators, training on OpenStreetMap tools is required.

Due to the characteristics of OpenStreetMap as an open database and the favelas being sensitive and conflicting spaces, it is necessary to consider some reflections regarding the privacy and security of the content made available on these spaces. In the absence of addresses or generic reference points, it is necessary to use abstract or personal references, such as the house of community leaders or known residents, or the indication of buildings that have unique characteristics. This information does not fit into the OpenStreetMap proposal, however, it is important for the use of favelas maps. Finally, using OpenStreetMap to create spatial information in favelas in conjunction with locals requires a lot of care, but this project has shown that its application is possible and can reduce the cartographic invisibility of vulnerable populations.

References

[1] Instituto Brasileiro de Geografia e Estatística (2010). *Censo demográfico 2010*. Instituto Brasileiro de Geografia e Estatística (IBGE), Rio de Janeiro, Brazil.

[2] Olthuis, K., Benni, J., Eichwede, K., & Zevenbergen, C. (2015). Slum Upgrading: Assessing the importance of location and a plea for a spatial approach. *Habitat International*, *50*, 270-288.

[3] Harris, T. M., Weiner, D., Warner, T. A., & Levin, R. (1995). Pursuing social goals through participatory geographic information systems. *Ground truth: the social implications of geographic information systems*, 196-222.

[4] National Center for Geographic Information and Analysis (1996). *GIS and society: the social implications of how people, space and environment are represented in GIS*. University of California, Santa Barbara, United States.

[5] Sieber, R. (2006). Public participation geographic information systems: A literature review and framework. *Annals of the association of American Geographers*, *96*(3), 491-507.

[6] Elwood, S. (2014). Straddling the fence: Critical GIS and the geoweb. *Progress in Human Geography*, 1-5.

How knowing the purpose of mapping changes the map and the mappers themselves

Patricia Solís[1,*]

[1] YouthMappers, Knowledge Exchange for Resilience Arizona State University, Tempe, AZ, United States; patricia.solis@asu.edu

* 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.

The opportunity presented through the YouthMappers network of chapters to engage university students in authentic, open humanitarian mapping raises important questions about how to guide the quality and productivity of volunteer spatial contributions while providing a valuable learning experience. It presents the unique chance to pique new mappers' interest, satisfaction, and confidence in spatial technologies in particular, and technology in general, as well as pique their interest in the people and places served by the humanitarian mapping projects. Our study explores the importance of sharing authentic contextual information about the purpose of the humanitarian mapping task [1].

Two groups of beginner mappers were given mapping tasks, with only one group being provided details on the purpose. Comparisons were made on their respective performance and changes in affective response to series of questions about technology, education, good citizenship, and empathy. Measures of the quantity and quality of spatial data produced, of their respective levels of interest, satisfaction, and confidence in technology, and affective responses before and after mapping show the relative effects of contextual information. Given our long-term aim of improving understanding about the experiences of new mappers who volunteer for humanitarian tasks, the results of this study are insightful about productivity, quality, and motivation [2] of students. Our findings indicate that purposeful, open humanitarian mapping [3] might help shift the difficult process of learning new technologies from a negative to positive affective experience, and potentially transfer authentic learning [4] activity focus from tools to context, an essential step toward achieving higher level learning objectives. These approaches may also even inspire students' future orientation to give back to society as productive, good citizens as well as spark their interest in geospatial technology careers.

With respect to mapping productivity and quality, knowing that the purpose of mapping is for humanitarian needs seems to make no difference in actual observed performance of beginners, except that some mappers with contextual information may introduce more types of errors. This finding suggests that a triage approach to training new mappers might be a novel strategy for organizers of humanitarian mapping campaigns to quickly assess the very different abilities of new volunteers, then pay greater attention to improve performance of those who need it most, to optimize overall production and quality.

Solís (2019). How knowing the purpose of mapping changes the map and the mappers themselves
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, 39-40. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387729

The fact that the performance perceptions of the informed group were much higher, yet their actual metrics showed little difference compared to the control group, it is imperative to seek ways to overcome what we term “The Do-Good” effect: Humanitarian mapping volunteers may believe they are doing well just because they are doing good. Strategies to combat this tendency should be sensitive to maintaining newcomer enthusiasm while paying attention to quality control through real-time validation and positive feedback.

Perhaps the clearest outcome from this study are the extremely strong results related to positive introduction of mapping technologies, even to students who are not inclined to be enthusiastic about technology in general. Merely performing the mapping activity helped all participants, including the control group, to better understand what mapping technologies are. When performed in the context of something as compelling as humanitarian mapping, the effect is even more profound. This finding suggests that humanitarian mapping may be one creative way to interest students in science and technology fields in general, whether or not they ultimately plan to focus on geospatial fields. Incorporating humanitarian mapping exercises in a variety of interdisciplinary experiences could strengthen educational goals of raising interest and motivation for STEM (Science, Technology, Engineering and Mathematics education).

These early results suggest that technical mapping activities offer a means to support affective learning toward good citizenship, giving back, and an empathic interest in the lives of others. For those who were provided contextual information about the humanitarian mapping task, they were significantly more likely to find it important to plan for being a good citizen while in college and for giving back to society. Mappers who were informed of the humanitarian purpose of their activity reduced the tendency to become generally less empathic from the stress of learning a new technology. Their interest in the feelings of other people may be an important place to start to leverage the growth of online mapping technologies to teach about human geography, place, and people within the framework of international humanitarian mapping.

Acknowledgements

The United States Agency for International Development supports the YouthMappers program through a grant from the USAID GeoCenter Award #AID-OAA-G-15-00007. Views expressed in this document do not necessarily reflect the policies and viewpoints of the agency or its employees. Founding partners are Texas Tech, George Washington University, and West Virginia University. Dedicated to the memory of James Eshun, Cape Coast University, Ghana.

References

[1] Solís, P., & DeLucia, P. (2019) Exploring the Impact of Contextual Information on Student Performance & Interest in Open Humanitarian Mapping. *Professional Geographer, 71*(3), 523-535.
[2] Budhathoki, N. R., & Haythornthwaite, C. (2012) Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap. *American Behavioral Scientist, 57*(5), 548-575.
[3] Solís, P., McCusker, B., Menkiti, N., Cowan, N., & Blevins, C. (2018) Engaging global youth in participatory spatial data creation for the UN Sustainable Development Goals: The case of open mapping for malaria prevention. *Applied Geography*, *98*, 143-55.
[4] Solís, P., Huynh, N., Carpenter, D., Adames de Newbill, M., & Ojeda, L. (2017). Using an authentic project based learning framework to support integrated geography education linked to standards and geospatial competencies. *Research in Geographic Education, 19*(20), 36-65.