International Cartographic Association
Association Cartographique Internationale

An AI-Based Approach To Identify Spatio-Temporal Contribution Patterns And Their Impact On OpenStreetMap Data Quality

Nathan Damas Antonio [a,b,*], Silvana Camboim [a], Alexander Zipf [b]

[a] *Department of Geomatics, Federal University of Paraná (UFPR) – nathandamas@ufpr.br, silvanacamboim@ufpr.br,*
[b] *Department of Geography, Heidelberg University - nathandamas@uni-heidelberg.de, zipf@uni-heidelberg.de*

* Corresponding author

Keywords: Volunteered Geographic Information, OpenStreetMap, Data Quality, Intrinsic Parameters**.**

Abstract:

This work enters the field of Geospatial Information Technology, targeting the lack of spatial data prevalent in developing regions. Voluntary Geographic Information (VGI) has become one of the major research topics in GIScience, more than fifteen years after the term was coined by Goodchild (2007). In many developing countries, where spatial data deficits are common at both large and small scales, Silva and Camboim (2020) suggest that VGI can mitigate the consequences of this scenario. Utilizing VGI via OpenStreetMap (OSM), the study aims to enhance urban planning efforts in Brazil.

OpenStreetMap, with over 7.8 million registered users and more than 16.7 billion edited features, is recognized as one of the world's largest and most popular platforms for collaborative data collection Teimoory, Abbaspour, and Chehreghan (2021). OSM has a significant advantage over commercial geodata providers because it adheres to the Open Data Commons Open Database License, which grants unrestricted data availability for a wide range of applications, promoting greater usability and flexibility. Researchers such as Brovelli and Zamboni (2018), Nasiri et al. (2018), and Haklay (2010) have focused their efforts on understanding the characteristics of OSM, particularly in terms of data quality and feasibility in integration processes. Understanding this is crucial for effectively using collaborative datasets like OSM.

ISO 19157 (ISO, 2013) is the prevailing global standard for assessing spatial data quality, including completeness, thematic accuracy, logical consistency, temporal quality, positional accuracy, and usability elements. In 2015, the Geographic Service Directorate in Brazil created the Technical Specification for Geospatial Data Quality Control (ET-CQDG). ET-CQDG lays out criteria based on the ISO 19157 standard for evaluating geospatial data quality. Furthermore, it presents various sophisticated sampling techniques to support the essential analyses. To complement these guidelines, in 2019, the Brazilian Institute of Geography and Statistics published the Technical Manual for Geospatial Data Quality Assessment, which has provided valuable standards and methods for evaluating the quality of this data.

While official agency maps exhibit uniformity due to professional expertise, VGI platforms do not guarantee this consistency. Consequently, global efforts focus on researching and analyzing VGI data quality parameters to assess their integration into official reference bases. Traditional measures often rely on comparisons with official data. However, Brovelli and Zamboni (2018) suggest using intrinsic parameters as an alternative approach to assess VGI quality. Mooney, Corcoran, and Winstanley (2010) support this, indicating intrinsic parameters' utility without comparative methods. Barron, Neis, and Zipf (2014) identified intrinsic indicators as feasible when quality is not strictly tied to official evaluation parameters. Grinberger et al. (2021) and Brückner et al. (2021) conducted mathematical modeling studies in OSM, analyzing contributions over time to detect global events and estimating the completeness of OSM retail stores in Germany, respectively.

Building on VGI, this research employs logistic regression models to analyze the completeness and growth dynamics of urban mapping within OSM. By collecting historical OSM data through the OSHOME API, the study applies logistic regression to model mapping activities' temporal evolution. This model identifies different mapping stages—initial growth, expansion, and saturation—providing a detailed urban mapping assessment.

The core objective is to use AI to systematically investigate spatio-temporal contribution patterns and their impact on OSM's geospatial information quality. This involves: (1) collecting OSM data for Brazilian cities from 2008 to 2023, (2) preprocessing the data using Python libraries like Pandas and Geopandas, (3) applying a four-parameter logistic regression model to the cumulative contribution history, and (4) extracting and analyzing logistic parameters to categorize cities based on mapping growth dynamics.

Abstracts of the International Cartographic Association, 7, 6, 2024.
European Cartographic Conference – EuroCarto 2024, 9–11 September 2024, TU Wien, Vienna, Austria.
https://doi.org/10.5194/ica-abs-7-6-2024 | © Author(s) 2024. CC BY 4.0 License.

The logistic regression model, described by Elias (2023), includes parameters defining the upper and lower asymptotes, the midpoint, and the curve's steepness. These parameters help understand the saturation level of mapping activities, crucial for urban planning. By visualizing logistic curves and categorizing cities, the study offers insights into urban mapping efforts' variability and maturity.

Focusing on intrinsic parameters, this study enhances the understanding of VGI data quality without relying solely on comparisons with official datasets. The expected outcomes will provide significant insights into VGI data's accuracy and utility, supporting sustainable urban development and urban planning efforts in Brazil. This research contributes to the global discourse on integrating VGI into official reference bases, particularly in advancing the United Nations' Sustainable Development Goals, specifically SDG 11—Sustainable Cities and Communities. According to Herfort et al. (2022), it aids urban planning, inclusive urbanization, public health, and environmental sustainability.

The methodology involves a robust AI approach, utilizing logistic regression to model mapping growth and leveraging the OSHOME API for comprehensive data collection. This innovative approach addresses spatial data deficits and provides a scalable framework for evaluating and improving VGI data quality on collaborative mapping platforms like OSM.

Acknowledgements

"This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001".

References

Barron, C., Neis, P., and Zipf, A., 2014. A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. Transactions in GIS, Vol. 18, No. 6, pp. 877–895.

Brovelli, M. A. and Zamboni, G., 2018. A New Method for the Assessment of Spatial Accuracy and Completeness of OpenStreetMap Building Footprints. ISPRS International Journal of Geo-Information, Vol. 7, No. 8, p. 289.

Brückner, J. et al., 2021. Assessing shop completeness in OpenStreetMap for two federal states in Germany. AGILE: GIScience Series, Vol. 2, pp. 1–7, June 4, 2021, online publication.

Elias, E. N. N. et al., 2023. Exploring spatio-temporal patterns of OpenStreetMap (OSM) contributions in heterogeneous urban areas. Boletim de Ciências Geodésicas, Vol. 29, p. e2023005, June 19, 2023.

Goodchild, M. F., 2007. Citizens as sensors: the world of volunteered geography. GeoJournal, Vol. 69, No. 4, pp. 211–221.

Grinberger, A. Y. et al., 2021. An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. Transactions in GIS, Vol. 25, No. 2, pp. 622–641.

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, Vol. 37, No. 4, pp. 682–703.

Herfort, B. et al., 2022. Investigating the digital divide in OpenStreetMap: spatio-temporal analysis of inequalities in global urban building completeness. Published on August 26, 2022, online publication.

ISO, 2013. ISO 19157:2013. Available at: https://www.iso.org/standard/32575.html. Access date: March 31, 2024.

Mooney, P., Corcoran, P., and Winstanley, A. C., 2010. Towards quality metrics for OpenStreetMap. Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS'10), November 2, 2010, New York, NY, USA.

Nasiri, A. et al., 2018. Improving the Quality of Citizen Contributed Geodata through Their Historical Contributions: The Case of the Road Network in OpenStreetMap. ISPRS International Journal of Geo-Information, Vol. 7, No. 7, p. 253.

Silva, L. S. L. and Camboim, S. P., 2020. Brazilian NSDI Ten Years Later: Current Overview, New Challenges and Propositions for National Topographic Mapping. Boletim de Ciências Geodésicas, Vol. 26, No. 4, p. e2020018.

Teimoory, N., Ali Abbaspour, R., and Chehreghan, A., 2021. Reliability extracted from the history file as an intrinsic indicator for assessing the quality of OpenStreetMap. Earth Science Informatics, Vol. 14, No. 3, pp. 1413–1432.

Witt, R., Loos, L., and Zipf, A., 2021. Analysing the Impact of Large Data Imports in OpenStreetMap. ISPRS International Journal of Geo-Information, Vol. 10, No. 8, p. 528.