Journal of Engineering Science *Vol. XXXI, no. 4 (2024), pp. 157 - 165*
Fascicle *Architecture, Civil Engineering and Environment* *ISSN 2587-3474*
Topic *Architecture, Urbanism and Cadaster* *eISSN 2587-3482*

https://doi.org/10.52326/jes.utm.2024.31(4).11
UDC 004.658.2:332(478-25)

ANALYSIS METHODOLOGY OF OPENSTREETMAP DATA COMPLETENESS (CASE STUDY: CHISINAU CITY)

Rodica Sîrbu *, ORCID: 0000-0001-7892-8747

Technical University of Moldova, 168 Ștefan cel Mare Blvd., Chisinau, Republic of Moldova
* Corresponding author: Rodica Sîrbu, *rodica.sirbu@icg.utm.md*

Received: 11. 26. 2024
Accepted: 12. 17. 2024

Abstract. In recent years, special attention has been paid to free and open geospatial data sources accessible to the whole world, which have benefits for all sectors of the national economy. Taking into account the interest in this topic, the article proposes a data completeness analysis for OpenStreetMap, that is based on crowd sourcing, compared to data sources available at the national level, such as statistical information from Public Institution Real Estate Cadastre registers or the Spatial Data Infrastructure of the Republic of Moldova (NSDI), Agency Geodesy, Cartography and Cadastre. Thus, a general methodology for completeness analysis framework is designed based on the characteristics of OSM data, since is a well-known standard. Secondly, an attempt was made to apply a quality assessment model for OSM data through three different quality elements: completeness, thematic accuracy and positional accuracy based on the visual and quantitative analyses taking into account the spatial data that already exists on national platforms. The Chisinau city was selected as a case study/area of interest, and the results of the presented research can subsequently be applied to the entire Republic of Moldova. Based on the analysis performed, it can conclude that spatial data sources, whether from official sources or those based on crowd sourcing, still need to be supplemented with both spatio-temporal and qualitative information.

Keywords: *volunteered geographic information, spatial data, Geographical Information Systems, OpenStreetMap, National Spatial Data Infrastructure.*

Rezumat. În ultimii ani s-a acordat o atenție deosebită surselor de date spațiale gratuite și deschise, accesibile tuturor, care au beneficii pentru toate sectoarele economiei naționale. Ținând cont de interesul deosebit față de această tematică, articolul propune o analiză a plenitudinii datelor spațiale de pe OpenStreetMap (OSM), care se bazează pe crowd sourcing, comparativ cu sursele de date disponibile la nivel național, precum informații statistice din registrele I.P. Cadastrul Bunurilor Imobile sau cele spațiale prezente Infrastructura de date a Republicii Moldova (INDS), Agenția Geodezie, Cartografie și Cadastru. Astfel, este concepută o metodologie generală pentru cadrul de analiză a plenitudinii pe baza caracteristicilor datelor OSM, deoarece sunt concepute în baza unui standard internațional. În al doilea rând, s-a încercat aplicarea unui model de evaluare a calității datelor OSM prin prisma a trei

elemente de calitate diferite: plenitudine, acuratețe tematică și acuratețe poziționlă pe baza analizelor vizuale și cantitative ținând cont de datele spațiale care există deja pe platformele naționale. Ca zonă interes a fost selectat orașul Chișinău, iar rezultatele cercetării prezentate pot fi ulterior aplicate pe întreg teritoriul Republicii Moldova. Pe baza analizei efectuate, se poate concluziona că sursele de date spațiale, fie din surse oficiale, fie cele bazate pe crowd sourcing, mai trebuie completate atât cu informații spațio-temporale, cât și calitative.

Cuvinte cheie: *informații geografice voluntare, date spațiale, Sisteme Informaționale Geografice, OpenStreetMap, Infrastructura națională de date spațiale*.

1. Introduction

For centuries, mapping of territory was carried out by state organizations. At first, military interests dominated, while in the 20th century, attention and application migrated to institutions responsible for land management, infrastructure planning or monitoring environmental components [1]. Meanwhile, Geographical Information Systems (GIS) have been used for decades, and currently they are no longer a tool used only by dedicated specialists, because the internet and web have significantly influenced the process of representing, searching, exploiting and use of geographic data. From the category of Web Mapping 2.0 applications, OSM remains to until today the best-known global geographic application project. OpenStreetMap Java Editor (JOSM) as a desktop application for editing OSM, being the most complex of the available editors today. The OSM project as a geographic application was initiated in 2004, for mapping the territory of Great Britain, and in 2006 the OpenStreetMap Foundation was founded for the development and distribution of free and open spatial data [2]. The geographical data in the OSM platform obtained according to the principle of "crowd-sourcing" is based on voluntary contributions of users to update the maps, which are subsequently accessible to the public as free and open data. At the same time, large companies such as Amazon, Apple, Facebook, Kaart and others contribute to improving the quality of OSM data [3]. Although Voluntary Geographic Information (VGI) has significantly improved the availability and use of geographic data, at the same time it has raised some reasonable concerns about their quality, reliability and overall value [4]. Although GIS experts have conducted research on the quality of OSM spatial data quality [5], it remains a current topic due to challenges related to the value of cartographic content, since the completeness of spatial information varies both at national and regional levels and also over time.

Currently, OSM data are most frequently used for urban infrastructure mapping [6,7], as they offer open access, global coverage, acceptable geo-precision, thematic and temporal plenitude [8] and are widely used by scientists and practitioners alike [9]. At the same time, national infrastructures are being developed all over the world with a multitude of reference data for different topics depending on state interest. But of course, the general users' interest in relation to spatial data is more focused on global sources because they are integrated into various common, well-known platforms with friendly and intuitive user interfaces, which is an additional reason to verify their quality.

2. Materials and methods

The initial materials used for the research were the specialized literature, the OSM and Geofabrik platforms, National Spatial Data Infrastructure (NSDI) of the Republic of Moldova, statistical data provided by the Public Institution Real Estate Cadastre. To assess the accuracy

of spatial data, a series of methods were used, from traditional ones, such as visual analysis, to the application of advanced processing tools in the specialized program. Some of the methods applied are strictly quantitative (Figure 1), while others are based only on visual observations taking into account the specifics of the information in the attribute tables or geometries. For all the processing described in this study, the specialized GIS software, free and open QGIS was used, as it represents a resource that can be accessible to all users.

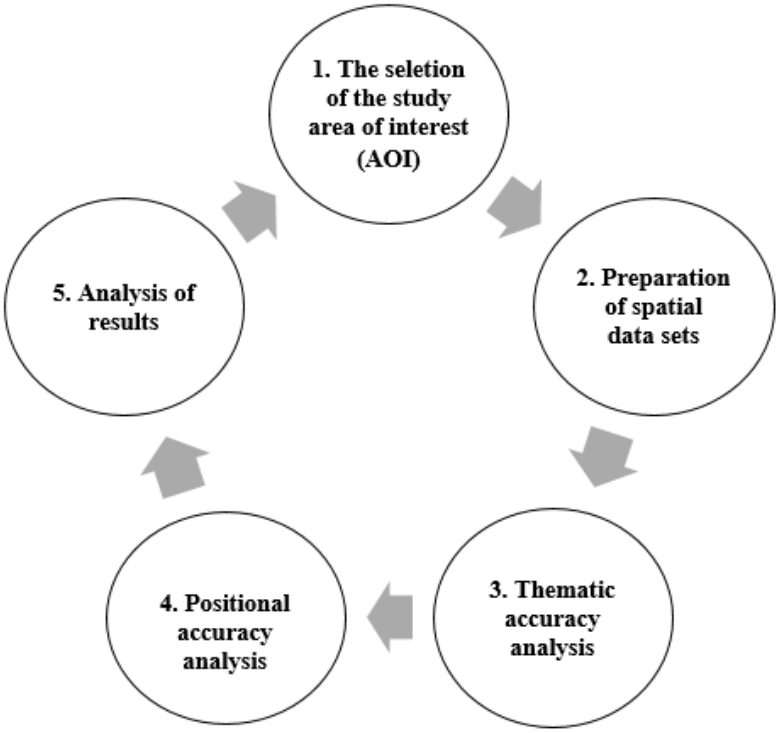

Figure 1. The methodology approached.

The Chisinau city, capital of the Republic of Moldova, was selected as the study area, which has an area of 122 km² [10] and a population of about 600 000 inhabitants [11]. Previous studies evaluated OSM in comparison with existing data to provide detailed information on selected cities [12], therefore, it is already an existing initial reference information. Since the method presented in the paper allows the assessment of positional accuracy and completeness of spatial data, therefore the analysis of previous literature focused mainly on these two aspects.

The process of preparing the spatial datasets had two objectives, namely firstly, the datasets must match each other in order to be able to compare them. Secondly, the datasets must be edited in such a way that they contain similar information, in order to facilitate the analysis procedure. As an element for the data analysis, the principle of completeness of information was used. Thus, the length and completeness of the name were analysed from the length or number of names of the tested dataset compared to the length or number of names of the reference dataset. The completeness of the spatial data was evaluated by comparing the OSM data with those accessed by the NSDI thematic geoportal and the data provided by the Public Institution Real Estate Cadastre. The thematic accuracy (accuracy of the name and classification) for the study area was analysed using the information from the attribute tables of the downloaded spatial data. The positional accuracy was determined by applying the query on OSM database or downloaded from Geofabrik [13], where a complete copy of the OSM database not older than one day is performed, and the instant download of

several existing datasets allows a more complex analysis, which enabled the selection and export of data categories for comparative analysis.

The OSM data were obtained using the QuickOSM plugin and through the query procedure were downloaded in shapefile format. For comparison, existing spatial data from the metadata of the NSDI platform of the Republic of Moldova were used. The GIS software used for the analysis was the free and open source QGIS, with the tools from the vector, plugin and geoprocessing tools on the menu bar.

3. Results

Large-scale projects, such as OSM, Wikimapia, GeoSam and others, provide access to spatial information for individual users, organizations and the scientific community without legal and economic restrictions, and among the most important benefits are minimal costs, accessibility, interoperability, data for less documented regions, etc. At the same time, these data are collected by users less experienced in GIS and consequently, the question arises regarding their quality. As is known, the methodology for analysing the quality of spatial data is very extensive, and through the existence of different standards and definitions, the following aspects can be considered [14]:

- ✓ Descent – refers to the way in which the dataset was collected and developed.
- ✓ Positional accuracy – evaluates the coordinates of an object (location).
- ✓ Attribute accuracy – objects in a geographic database are represented not only by their geometric shape but also by additional attributes, this measure evaluates how correct these values are.
- ✓ Logical consistency – topologically correctness and relationships that are encoded in the database.
- ✓ Completeness – coverage of objects existing in the real world.
- ✓ Semantic accuracy – links the way the object is integrated and represented in the database to its meaning and how it should be interpreted.
- ✓ Temporal quality – this is a measure of the validity of changes in the database in relation to real world changes and the rate of updates.

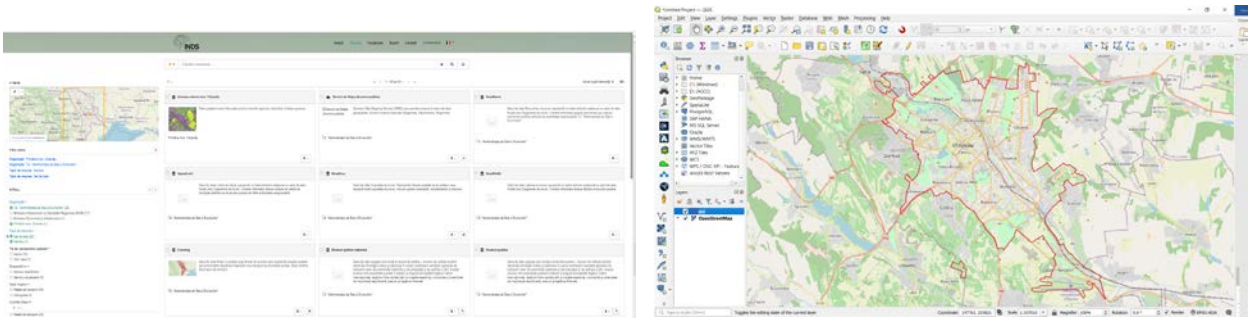

Figure 2. Selection of initial spatial data sets and delimitation of the AOI.

Thus, in order to proceed with the data analysis, existing metadata were identified using the filtering tool [15] through specific key words, identifying the required services or spatial data and as well the study area was delimited (Figure 2).

The next step was to install the QGIS plugin - QuickOSM, which made it easy to download and manage OSM data directly from the specialized software graphic user interface, without Overpass Turbo or other intermediary platforms. QuickOSM is a very useful tool for quickly downloading OSM data based on custom filters (Figure 3).

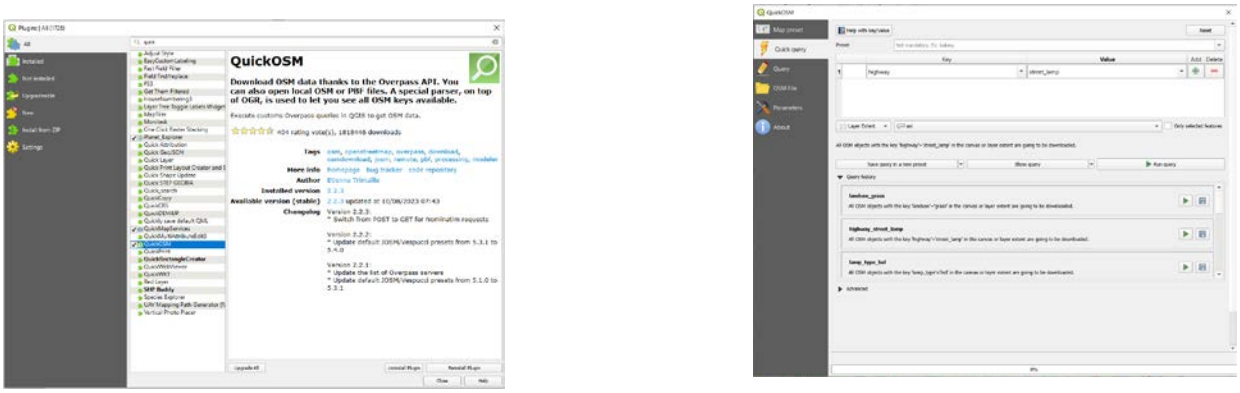

Figure 3. Installing the QuickOSM plugin and applying queries to OSM database.

As a result of the procedure for querying and downloading data from OSM (for example, street lighting), an information layer in vector format was obtained. To verify this data, other existing information layers at the national level were selected. For comparison, the metadata provided by the Chisinau City Hall and the metadata for the Linear Maps (Web Map Service (WMS) and Web Feature Service (WFS) allow viewing the base map - Base Map) 2017, for the territory of the Republic of Moldova, were selected. The linear map has approximately 90 layers with information about: administrative units, landscape, hydrography, buildings, roads, railways, utilities etc. The collection of objects was carried out based on orthoimages from 2016) which provide the possibility of accessing information through the WFS and WMS network services through specialized GIS programs.

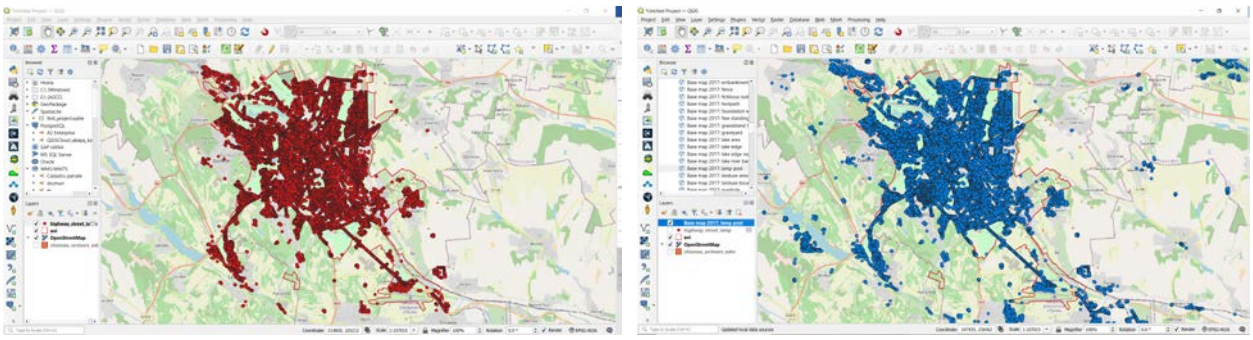

Figure 4. Analysis of SPATIAL datasets from OSM and Linear Maps.

In the city of Chisinau, sectors were identified where the OSM data (Figure 4) with greater coverage have a higher completeness than the official data existing on the NSDI or Geodata platform, or the State Geographic Information System at the local level for local public administration authorities (SIGSNL).

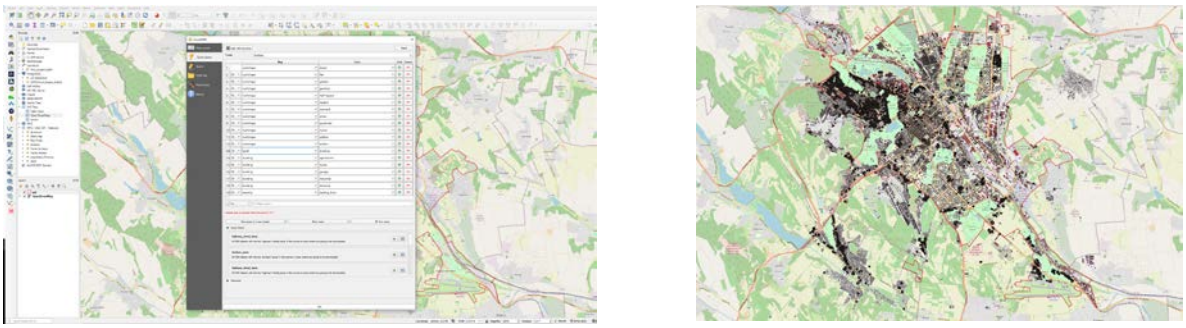

Figure 5. Query procedure for OSM data.

At the same time, another situation is attested in the analysis of residential buildings, using the same plugin, the query was performed based on the keyword constructions and the information referring to the residential buildings existing on the OSM was extracted for the AOI.

Analysing the obtained results, a discrepancy in the consistency of spatial data for different sectors of the Chisinau city was identified, namely from the OSM query (Figure 5) regarding constructions, where a total of 40954 units was obtained, and from the database of the P. I. Real Estate Cadastre - 3668 units. Thus, comparing the information provided by the relevant institution with the column in the attribute table building: levels and height, it was deduced that the information on the OSM requires updating, because the basic parameters for residential blocks: height (1085 units) and number of levels (17 units) are indicated at an insignificant number, while there is much more information (Figure 6).

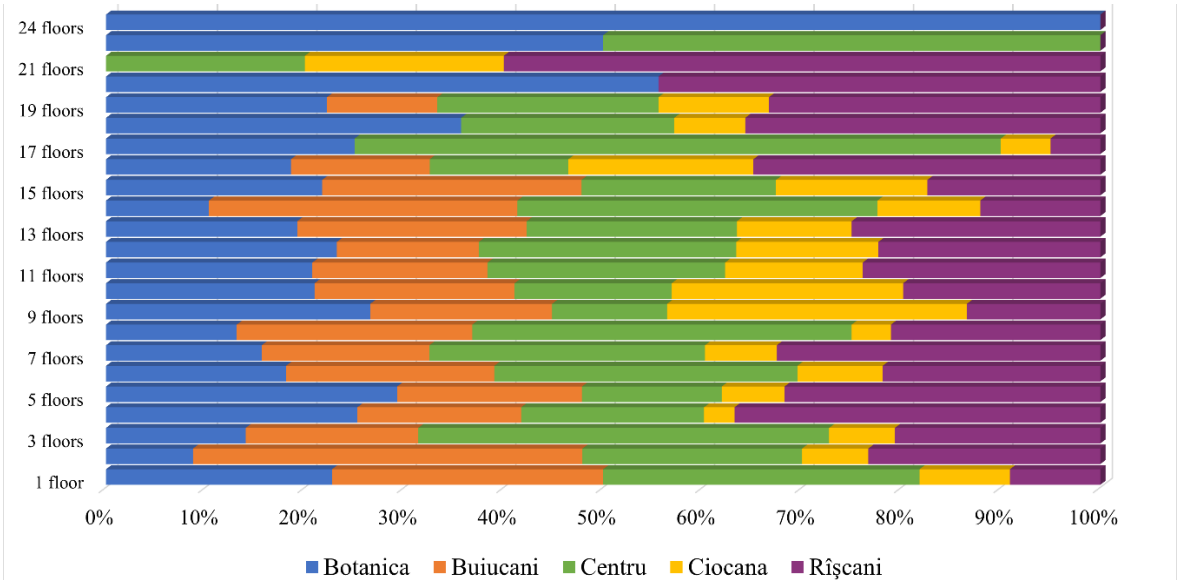

Figure 6. Analysis of residential buildings distribution in the Chisinau city [16].

Also, visually, by using the orthophoto image accessed through the visualization service (WMS) and by applying the Map Swipe tool plugin (Figure 7), entire sectors were identified where buildings are not represented on OSM. At the same time, by applying the Street View tool, it was established that many buildings are in reality private houses.

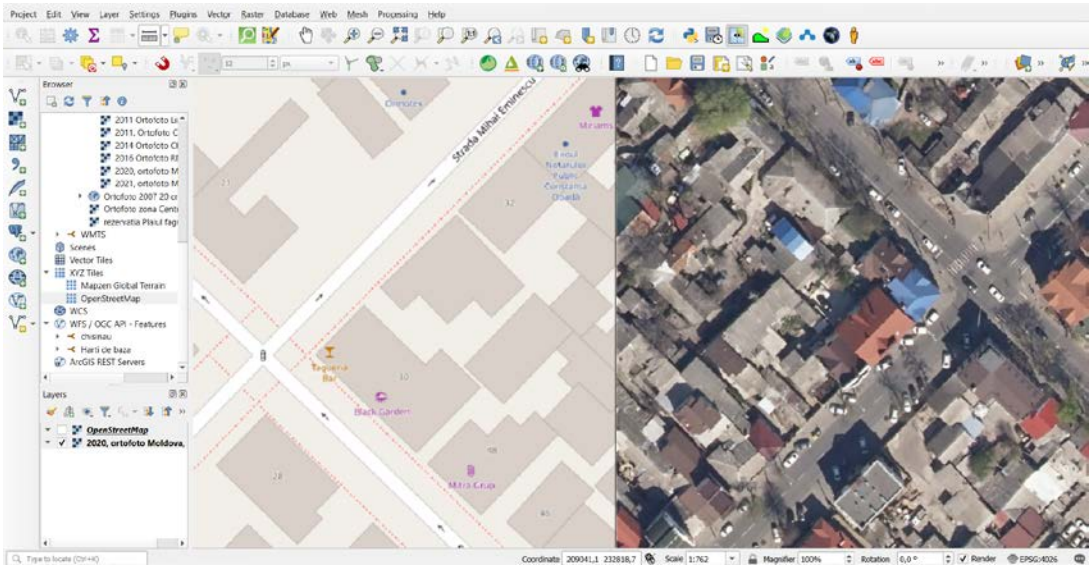

Figure 7. Applying the MapSwipeTool tool for spatial data comparison.

Another analysis was performed to assess the completeness of the data on road infrastructure. Thus, using the same plugin for downloading data from OSM, it was visually observed that the spatial data is much more qualitative, considering that this project has a great application for mobility and identifying the most optimal traffic directions. Also, the analysis of the availability of spatial data was performed with reference to the area/parking spaces available. Thus, for comparison, the data provided by the Chisinau City Hall (General Directorate of Architecture, Urbanism and Land Relations), through the download network services (WFS) and the data from OSM by applying the query using keywords, were used.

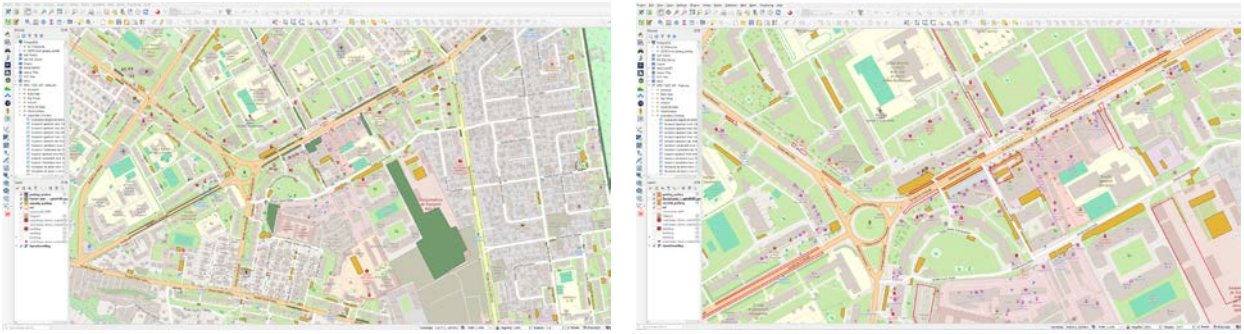

Figure 8. Comparison of data referring to parking available via download from OSM and SIGSNL download network services.

Following the visual analysis of the information, it was determined that it complements each other, and for a more detailed analysis of the current situation, also more layers and information sources should be used (Figure 8). The comparison with the data provided by the P. I. Real Estate Cadastre could not be carried out because from the available registers only information can be extracted referring to the garages within the residential blocks, the total number of which for the entire city of Chisinau is 3440, and the underground parking lots, which represent 12821 in number. Another analysis was carried out with reference to the spatial data about roads and access roads, and the initial information for the analysis was downloaded by Geofabrik and subsequently cropped for the area of interest (Figure 9).

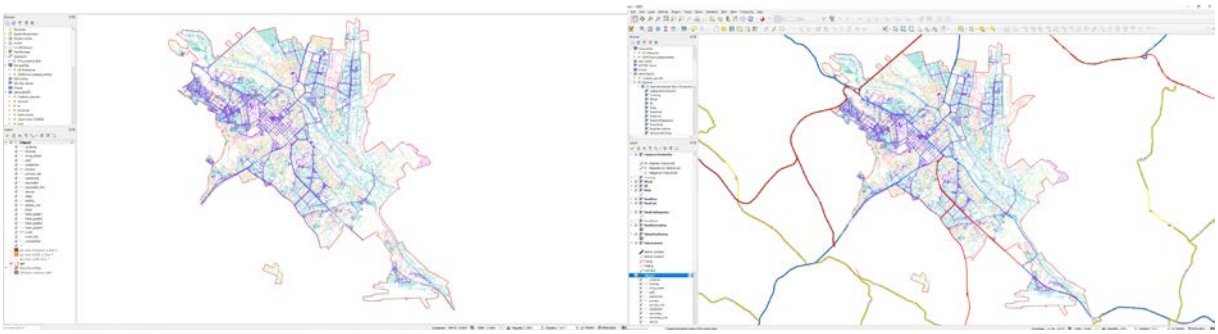

Figure 9. Comparison of data with reference to roads, OSM and the WMS service of the State Road Administration.

The comparison was made with the national data provided through visualization services by the State Road Administration, namely the spatial data set containing public roads - public utility roads intended for road and pedestrian traffic in order to meet the general road transport requirements of the national economy and the population and for the defense of the country. These roads represent public property of the state and ensure the main international road connections, the connection between the capital of the country and the cities-residences, municipalities and objects of republican importance, as well as the

connection between them [17] and through WMS visualization services - National public roads (Magistral, Republican and Regional ones).

4. Discussions

Accurate and up-to-date cartographic products are now an essential resource for a wide range of applications, education, science but also for the development of society in general. Producing and updating spatial data in a constantly changing world remains a major scientific and practical challenge especially for public institutions, where the process of producing and updating spatial data is currently very complex and bureaucratic. The involvement of citizens in the collection of geographic data has greatly influenced the development and use of maps and related data, having an impact on the performance of every routine activity. Digital resources such as Google Maps, Google Earth, Bing Maps and even maps that are generated by users through projects such as OSM are widely used by various categories of users. At the same time, verifying the correctness of these data is equally important, and the procedure for verifying the completeness and accuracy of these data remains an open problem that will continue to require a complex approach.

It has been shown that OSM is now widely used for applications such as web maps and navigation services, and OSM data has been used in areas such as urban planning [18], monitoring of sustainable development goals [19], disaster management [20], public health [21], to support crisis response during the COVID-19 pandemic [22], etc. Considering the very vast areas where spatial data can be used, the process of permanent analysis of consistency, accuracy is very important, especially for urban areas where a larger number of geographic information users are concentrated.

5. Conclusions

The development trend of VGI collection will become even more integrated with contemporary technologies, such as artificial intelligence, augmented reality, the Internet of Things (IoT) and gamification. Free and open global spatial data are valuable resources in the teaching-learning process for any level, but also with particular application in the economic development and research process in different fields. Thus, this evolution will transform open-source platforms into a fundamental pillar for various industries, increasingly involving communities and providing a collaborative framework for solving local and global problems.

The purpose of this research was to assess the completeness of VGI spatial data and in particular, to assess the completeness of data from the OSM project. Although there are previous researches on the quality and completeness of OSM data, this study is the first attempt carried out for the Chisinau city, Republic of Moldova. The methodology used consisted of comparing the completeness elements of OSM data with the NSDI Republic of Moldova spatial data and different sources and registers with data related to the topic of the presented research. In general, the results showed that OSM has accurate and up-to-date data, which indicates the usefulness of the data for developing projects in various fields of activity.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Heipke, Ch. Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing* 2010, 65(6), pp. 550-557.
2. Wang, M.; Li, Q.; Hu, Q.; Zhou, M. Quality analysis of open street map data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 2013, pp.155-158.

3. Anderson, J.; Dipto S.; Leysia P. Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information* 2019, 8(5), 232.
4. Flanagin, A.; Metzger, M. The Credibility of Volunteered Geographic Information. *GeoJournal* 2008, 72, pp. 137-148.
5. Borkowska, S.; Pokonieczny, K. Analysis of OpenStreetMap Data Quality for Selected Counties in Poland in Terms of Sustainable Development. *Sustainability* 2022, 14, 3728.
6. Haklay, M. How Good Is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design* 2010, 37, pp. 682–703.
7. Fan, H.; Zipf, A.; Fu, Q.; Neis, P. Quality Assessment for Building Footprints Data on OpenStreetMap. *Int. J. Geogr. Inf. Sci.* 2014, 28, pp. 700-719.
8. Raifer, M.; Troilo, R.; Kowatsch, F.; Auer, M.; Loos, L.; Marx, S.; Przybill, K.; Fendrich, S.; Mocnik, F.-B.; Zipf, A. OSHDB: A Framework for Spatio-Temporal Analysis of OpenStreetMap History Data. *Open Geospat. Data Softw. Stand.* 2019, 4, 3.
9. Herfort, B.; Lautenbach, S.; Porto de Albuquerque, J.; Anderson, J.; Zipf, A. The Evolution of Humanitarian Mapping within the OpenStreetMap Community. *Sci. Rep.* 2021, 11, 3037.
10. National Geospatial Data Fund. Available online: https://geoportal.md/ (accessed on 25.11.2024).
11. Cara, O. Chisinau in Figures, Statistical Yearbook. National Bureau of Statistics of the Republic of Moldova, Chișinău, 2023, 116 p.
12. Brovelli, M.; Zamboni, G. A New Method for the Assessment of Spatial Accuracy and Completeness of OpenStreetMap Building Footprints. *ISPRS Int. J. Geo-Inform* 2018, 7, 289.
13. Geofabrik. Available online: https://download.geofabrik.de/ (accessed on 20.11.2024).
14. Van, O. Spatial data quality: from description to application. PhD Thesis, Wageningen Universiteit, Wageningen, 2006.
15. National Spatial Data Infrastructure of the Republic of Moldova. Available online: https://geoportalinds.gov.md (accessed on 29.11.2024).
16. Public Institution Real Estate Cadastre. Available online: https://ipcbi.gov.md/ro/content/e-cadastru (accessed on 29.11.2024).
17. Metadata. Available online: https://geoportalinds.gov.md (accessed on 29.11.2024).
18. Milojevic-Dupont, N.; Hans, N.; Kaack, L.H.; Zumwald, M.; Andrieux, F.; de Barros Soares, D. Learning from urban form to predict building heights. *PLoS ONE* 2020, 15, pp. 1–22.
19. Van Den Hoek, J.; Friedrich, H. K.; Ballasiotes, A. Development after Displacement: Evaluating the Utility of OpenStreetMap Data for Monitoring Sustainable Development Goal Progress in Refugee Settlements. *ISPRS Int. J. Geo-Inform.* 2021, 10, 153.
20. Scholz, S.; Knight, P. Volunteered Geographic Information for Disaster Risk Reduction-The Missing Maps Approach and Its Potential within the Red Cross and Red Crescent Movement. *Remote Sens.* 2018, 10, 1239.
21. Pinho, M.; Flueckiger, B.; Valentin, A.; Kasdagli, M.-I.; Kyriakou, K.; Lakerveld, J.; Mackenbach, J-D.; Joline W.J. Beulens, J.W.J.; Kees de Hoogh. The quality of OpenStreetMap food-related point-of-interest data for use in epidemiological research. *Health & Place* 2023, 83, 103075.
22. Mooney, P.; Grinberger Y.; Grinberger Y. OpenStreetMap Data Use Cases During the Early Months of the COVID-19 Pandemic. In: *COVID-19 Pandemic, Geospatial Information, and Community Resilience*, CRC Press, 2021, 16 p.

Citation: Sîrbu, R. Analysis methodology of openstreetmap data completeness (case study: Chisinau city). *Journal of Engineering Science* 2024, XXXI (4), pp. 157-165. https://doi.org/10.52326/jes.utm.2024.31(3).01.

Submission of manuscripts: jes@meridian.utm.md