OpenStreetMap data quality assessment according to ISO 19157-1:2023 for Slovenia and Bosnia and Herzegovina

Adna Vukalić [a,*], Mihaela Triglav Čekada [a,b], Dušan Petrovič [a,*]

[a] University of Ljubljana, Faculty of Civil and Geodetic Engineering, adnavukalic12@gmail.com, Mihaela.Triglav-Cekada@fgg.uni-lj.si, Dusan.Petrovic@fgg.uni-lj.si
[b] Geodetic Institute of Slovenia

* Corresponding author

Keywords: quality assessment, quality elements, OpenStreetMap, ISO 19157-1:2023, Slovenia, Bosnia and Herzegovina

Abstract:

Data quality is one of the key factors that determine the choice of dataset for a particular purpose. OpenStreetMap (OSM) is the most widely used and the most complete database of voluntarily collected geographic information. That geographic information is mostly collected by volunteers, often without specific knowledge of geoinformatics (VGI). Their quality is very heterogenous and is often cited as a major obstacle to their wider use, mainly due to the use of different methods in their creation and the fact that the data can be created by anyone. So, it is not surprising that assessing the quality of these data is the subject of a great number of studies. The quality assessment must be objectively determined, therefore, based on this requirement, we have assessed the quality of OpenStreetMap data according to the quality elements and according to the procedures as prescribed by ISO 19157-1:2023.

We analysed three types of test areas in Bosnia and Herzegovina and Slovenia, representing urban, rural and high mountain areas. For each type of test area, we analysed two cases for both countries. Each test area is defined in a square with an area of 9 km², which means that we analysed 54 km² in the Federation of Bosnia and Herzegovina and 54 km² in Slovenia. The test areas were chosen in such a way that two test areas of the same type in the countries analysed are as similar as possible in terms of area, population and development. When defining the test areas within cities, we assumed that there would be no major changes in the inner city, so in these cases we focused on the periphery of the city. The rural test areas consisted of one smaller and one larger village, while the high mountain test areas consisted of a mountain peak and its surroundings. The choice of test areas was also influenced by the availability of up-to-date control data in both countries. The reference source used in the OpenStreetMap data quality analysis for the selected test areas in Slovenia and Bosnia and Herzegovina is the most recent national orthophoto image. We also used LiDAR data, building cadastre data, public utility infrastructure data, and records of land use of agricultural and forest land.

The analysis focused on the quality assessment of the following quality elements - completeness, positional accuracy and thematic accuracy. The aim was to find out whether the quality of the data depends on the type of area and whether it differs between the two countries. Based on a detailed analysis, we found that the quality of the data varies according to the type of test area (urban, rural, high mountain) for both the areas in Bosnia and Herzegovina and Slovenia, with even greater differences observed when comparing the same type of test area in the two countries analysed. A key finding is that the quality of the data in OpenStreetMap is significantly dependent on the involvement of volunteers in a given area, and only then on the type of test area.

If we highlight the results of the completeness of buildings and transportation networks (roads, tracks, and paths) for both analysed countries, we can observe the following. The completeness of building data is comparable between the countries, with no drastic deviations. In Slovenia, high mountain test areas show 100 % completeness, which can be attributed to the high visitation rates of both tested areas. The analysed areas cover the peaks and their immediate surroundings, where most of the buildings with specific thematic content, mainly mountain huts, are located. These are of great importance for these areas, which contributes to the high completeness of the data covered. This is followed by urban areas with 92.6 % and rural areas with 76.2 % completeness. The results for the urban test areas are satisfactory, especially because we focused on the peripheries, when selecting areas for city analysis, which are more susceptible to changes. We have also achieved outstanding results in the rural test areas, particularly because our selection included both a more developed and a less developed village. In Bosnia and Herzegovina, urban test areas show slightly lower completeness compared to rural areas, likely due to a large number of newly constructed buildings. The completeness of high mountain areas is lower, as one of the test areas is less known compared to all the others. Buildings not covered in the OpenStreetMap portal in urban test areas are mostly newly

built residential structures, while the missing data in rural test areas are most often isolated residential and agricultural buildings.
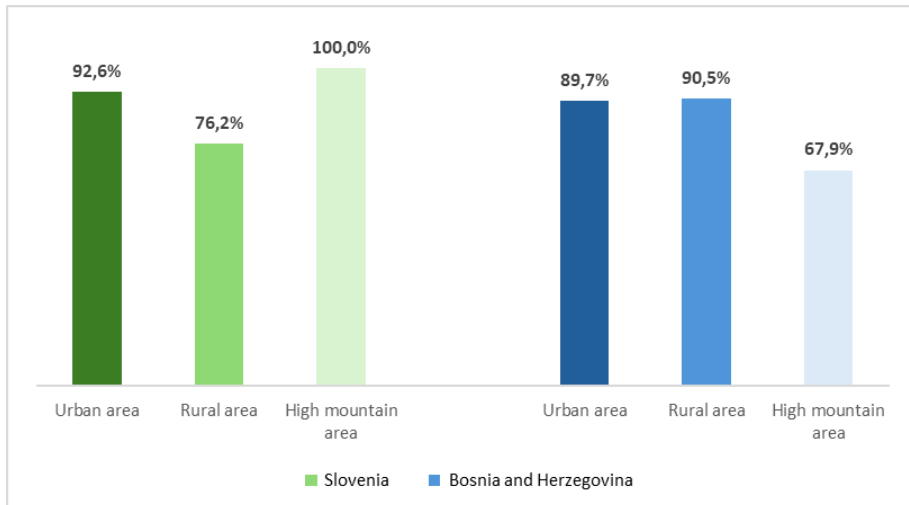

Figure 1: Completeness of buildings

Among the different types of test areas (urban, rural, high mountain) and compared to the other types of transport networks (tracks and paths), roads were the most complete in both countries. Tracks completeness is relatively equally high in urban and rural test areas in both countries. Paths completeness is very high in most of the test areas in Slovenia. In Bosnia and Herzegovina, path completeness is markedly high in the high mountain test areas, which is understandable given the importance of paths in these regions, while completeness in urban and rural areas is less than 50 %.
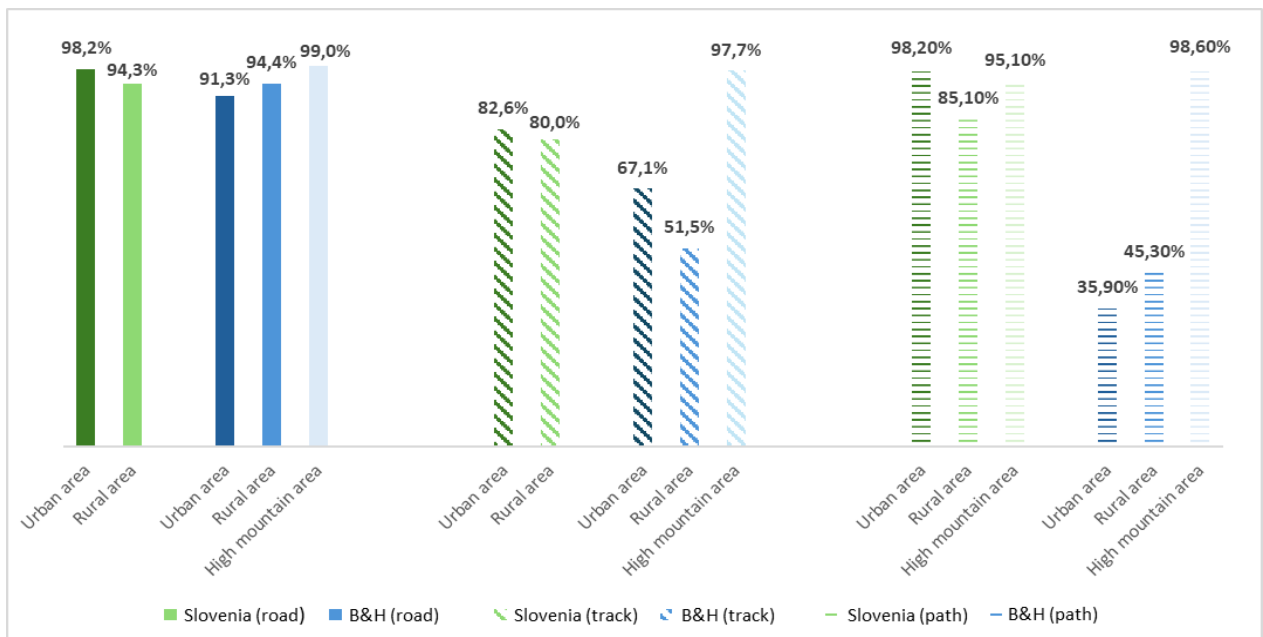

Figure 2: Completeness of roads, tracks and paths

It is important to note that the differences in results for urban, rural and high mountain areas are not as pronounced as in some other studies. This is because our test areas do not include megalopolises, unpopulated rural areas or inaccessible high mountain areas, where we would expect much larger differences in OpenStreetMap data quality.

Acknowledgements