

# AUTHORITATIVE CARTOGRAPHY IN BRAZIL AND COLLABORATIVE MAPPING PLATFORMS: CHALLENGES AND PROPOSALS FOR DATA INTEGRATION

## *Cartografia de referência do Brasil e dados provenientes de plataformas de mapeamento colaborativo: desafios e propostas para integração*

Leonardo Scharth Loureiro Silva<sup>1</sup> - ORCID: 0000-0002-9226-6775

Silvana Philippi Camboim<sup>1</sup> - ORCID: 0000-0003-3557-5341

<sup>1</sup> Universidade Federal do Paraná, Programa de Pós-Graduação em Ciências Geodésicas, Curitiba - Paraná, Brasil.  
E-mail: scharth.leo@gmail.com, silvanacamboim@gmail.com

Received in 30<sup>th</sup> June 2020

Accepted in 29<sup>th</sup> November 2020

### Abstract:

Brazil has a large area with missing or outdated mapping on the largest scales of its authoritative mapping. The use of data from collaborative mapping platforms appears as an alternative that may contribute to minimizing this problem, either by updating or completing the mapping coverage in Brazil, as proposed or performed by some National Mapping Agencies abroad. The present work aims to analyze a methodology to provide accurate and documented integration of volunteered geographic information and the Brazilian authoritative mapping. The proposal starts with the semantic compatibility between the conceptual models adopted in both official cartography and OpenStreetMap platform. The research continues with the identification of object classes with the most significant potential for integration. Finally, we developed some experiments to evaluate and validate the OSM data integration process in a 1:25,000 scale cartographic database. Even in regions with a recent mapping, the results of the preliminary assessment indicate the potential for an increase of about 52% and 16% of features in the 'road system' category, which suggests a very promising method for use in areas with missing or outdated mapping, and its applicability to other categories.

**Keywords:** VGI; Authoritative mapping; OpenStreetMap; Cartographic updating; Semantic compatibility; NMA.

### Resumo:

O Brasil tem uma grande área com mapeamento desatualizado ou sem cobertura cartográfica nas maiores escalas de seu mapeamento de referência. O uso de dados de plataformas colaborativas de mapeamento figura como uma alternativa que pode contribuir para minimizar esse problema, seja através da atualização seja na complementação da cobertura do mapeamento no Brasil, conforme proposto ou realizado por algumas Agências Nacionais de Mapeamento no exterior. O presente trabalho tem como objetivo analisar uma metodologia para fornecer uma integração acurada e documentada das informações geográficas voluntárias ao mapeamento de referência brasileiro. A proposta tem início na compatibilização semântica entre os modelos conceituais adotados na cartografia oficial e

**How to cite this article:** SILVA, L.S.L.; CAMBOIM, S.P. Authoritative cartography in Brazil and collaborative mapping platforms: challenges and proposals for data integration. *Bulletin of Geodetic Sciences*. 27(spe): e2021003, 2021.



This content is licensed under a Creative Commons Attribution 4.0 International License.

na plataforma OpenStreetMap, passa pela identificação das classes de objetos com maior potencial para a integração e, finalmente, apresenta experimentos iniciais para avaliar e validar o processo de integração de dados OSM em um banco de dados cartográficos na escala 1:25.000. Mesmo em regiões com mapeamento recente, os resultados da avaliação preliminar indicam potencial para incremento de 52% e 16% nas feições da categoria 'sistema rodoviário', o que sugere um método promissor para uso em áreas com mapeamento desatualizado ou sem mapeamento, além de sua aplicabilidade a outras categorias.

**Palavras-chave:** VGI; Mapeamento de referência; OpenStreetMap; Atualização cartográfica; Compatibilidade semântica; Agências Nacionais de Mapeamento.

## 1. Introduction

More and more, human activities have become dependent and are optimized by the use of geospatial information (GI), which results in increasing diversity in the contexts of GI use (Harding et al. 2009). Both in daily personal activities and in the course of projects in private organizations, in the planning and operations carried out in public and governmental institutions as well as in scientific research, GI is a fundamental and strategic element that must be considered in decision-making.

Naturally, in the geodetic sciences this importance is also observed. The existence of accurate and updated authoritative cartographic bases, in a scale compatible with the specific uses, means an indispensable framework for understanding the reality of the region of interest of any project, action or research. Thus, these bases are the essential input for spatial analysis and the most varied studies about activities or facts that occurred in a given geographic space. Besides, the availability of historical cartographic data allows spatio-temporal analysis that can result in evaluations of the dynamics that have happened in a region.

However, Brazil has a historical deficiency in the coverage of cartographic data in its territory, noticeable in the larger scales established for the national topographic mapping. According to Silva and Camboim (2020), the 1:25,000 scale, for example, has hugely limited coverage (about 5.5% of the country) incorporated into the National Cartographic System. The intermediate scales, 1:50,000 and 1:100,000, when combined, cover 94% of the country, but on the other hand, mostly produced more than 40 years ago. Only for the scales 1:250,000 and 1:1,000,000, the country has complete and updated cartographic coverage, entirely produced following established norms and standards, integrated into a geographic database and part of a continuous updating program.

Despite Brazilian National Mapping Agencies increased the production of mapping at 1:25,000, 1:50,000 and 1:100,000 scales, in the recent years, Brazil still has a large area that needs cartographic updating and many others without mapping at the largest scales. In addition, its territory has regions with very heterogeneous characteristics, which requires specific planning for different types of mapping projects. To change this scenario, political and economic arrangements are necessary, with significant and permanent funding as well as a considerable number of qualified staff to be involved in the production.

Because of this scenario with many problems, a feasible alternative to the limitations of this current reality is the use of data from collaborative mapping platforms. Goodchild (2007) defined the term Volunteered Geographic Information (VGI), which can be understood in a simplified view, as a set of geographical data generated by people or organizations not specialized in geospatial data production.

Other important factors that stand out as positive points are the availability and widespread use of the internet through the use of mobile technologies increasingly accessible to citizens. Thus, ubiquitous computing in parallel with the trend of mobility for several daily activities can make ordinary citizens become a key player in Cartography, as an active and permanent provider of VGI. Furthermore, this new "data source", due to its diffusion, its spread

everywhere and for its terrific speed in updating geospatial information, it can mean an important element of communication of the areas that are most in need for the NMA efforts.

The use of geographic information based on voluntary contributions has experienced some successful approaches worldwide. Research by Olteanu-Raimond et al. (2017b) presents the most usual forms of communication or interaction between users and European mapping agencies such as alerts about errors, new data inclusion, reporting of local names, change detection and image interpretation.

Some positive results were verified, for instance, in the Netherlands, Finland and Switzerland agencies. In these NMAs, error alert tools or other applications to send volunteered information were developed, which after analysis and validation have resulted in changes in the organizations' databases (Olteanu-Raimond et al. 2017a; Olteanu-Raimond et al. 2017b). Also, as a successful example, USGS, the North American NMA, has an auspicious crowdsourcing project, TNMCorps. This project counts with a numerous network of volunteers who collect and edit a defined list of points of interest for the organization. It has already reached the expressive mark of 400,000 points, and a large part of these points has been validated and integrated into the US national topographic database (USGS 2019; Korris, Niknami and McCartney 2017).

Some reports, like those about Portugal, add challenging situations that can be understood as lessons learned from implementation in other countries (Olteanu-Raimond et al. 2017b).

Several authors (Coleman 2012; Sester et al. 2014; Basiouka, Potsiou and Bakogiannis 2015; Olteanu-Raimond et al. 2017b), over the past number of years, have been evaluating experiences, initiatives and potentialities to integrate VGI with authoritative mapping as well as the use of VGI in official mapping projects. There are also studies dealing with the combination of social media geographic information (SMGI) with authoritative data (Albuquerque et al. 2015; Massa and Campagna 2016). These findings show that geospatial data producers can use all these source of data, especially in areas with outdated information or missing data.

In the same approach, there is a large variety of studies involving the quality assessment by comparing OSM data with authoritative datasets in sundry countries, such as United Kingdom (Haklay 2010), France (Girres and Touya 2010), Austria (Graser, Straub and Dragaschnig 2013), Germany (Hecht, Kunze and Hahmann 2013; Dorn, Törnros and Zipf 2015), Iran (Forghani and Delavar 2014) and Italy (Brovelli et al. 2016; Brovelli et al. 2017; Fonte et al. 2018). Hence, in particular, the assessment of positional accuracy and completeness of VGI has been widely studied in the last years.

Among the several works mentioned, two of them stand out as the precursors in the comparative evaluations between OpenStreetMap and authoritative mapping. Haklay (2010) analyzed the quality of OSM datasets in the UK by evaluating positional and attribute accuracy, completeness and consistency. He used the buffer zone method to compare linear features and calculated the overlapping of homologous lines. Girres and Touya (2010) performed a more extensive set of evaluations in eight quality elements for OSM data in several France regions. In the assessment of linear features, they used Hausdorff distance and the average distance.

This paper's proposal contains some aspects that overlap with previous works, such as the importance of making semantic compatibility between the official and VGI models and using the buffer zone for comparison between homologous lines. On the other hand, the most remarkable difference between the previous and the present work is that, in this article, OSM data are not compared with another dataset considered the ground truth. The core idea is to identify the data represented in OSM that are not present in the reference database. These are treated as potential data to be integrated into the official mapping, important information in countries with inadequate authoritative mapping coverage, such as Brazil and other developing countries. Another particular aspect is that most of the studies cited were focused on object classes or features compatible with representation on large scales, especially street or road networks and buildings. This is another significant difference in the present work, whose comparison datasets are in the context of smaller scale maps, since there is not much availability of the scale ranges from 1:25,000 and larger in Brazil.

Interoperability issues (Al-Bakri and Fairbairn 2012) and semantic compatibility (Machado, 2020) have also been investigated to the handling of the heterogeneity between authoritative and collaborative data models. This is a significant step in the integration of VGI with official mapping and, consequently, Spatial Data Infrastructures (SDIs). Al-Bakri and Fairbairn (2012) affirm that the most crucial problem of heterogeneity in spatial datasets comes from the lack of semantic and structural correspondences of the data from distinct sources.

In general, authority datasets have well-established, documented and rigid semantic characteristics while collaborative platforms have a free and flexible aspect. Bordogna et al. (2016) describe the architecture of an SDI for creating, mapping and managing full interoperable VGI. It is vital to solving these issues to make SDIs more modern and closer to the needs of users. Mooney and Corcoran (2011) believe that the generation of knowledge from multiple sources, including VGI, can play a crucial role in building a more accessible SDI.

SDI was initially conceived as an environment to support the coordination, integration, exchange and sharing of data and spatial metadata that mainly served the public sector. Harvey and Tulloch (2006) describe in their research that as the SDI concept and implementation matured, issues that could impair SDI were identified. Despite the availability of technologies and concepts, political and governance actions became the biggest obstacles. Thus, as over the years, conceptual changes in SDI have been shaped. These changes have resulted in new generations of SDI with the opening of new opportunities. Johnson (2017) describes potential spatial data management models that meet the philosophy of services based on open data and integration of VGI. In this sense, the author includes key areas of research such as the development and implementation of user-editing interfaces and evaluation of the data received by government through any type of data curation. Finally, Johnson (2017) suggests that the sustainability of this strategy aimed at user participation must rely on both the continued engagement of volunteers and continued support by government.

Borba et al. (2015) state that in the context of SDIs, applying open initiatives means removing social, technical, technological, financial and legal barriers. Moreover, open initiatives encourage and facilitate the sharing and updating of information. Coetzee et al. (2013) reinforce that crowdsourcing and mobile technologies are having an evolutionary impact on SDIs and draw attention to the need is evident for further consideration of the future roles of standards and ontologies in assuring the integration of data sources for cartographic analysis and presentation. Borba et al. (2015) propose a contemporary infrastructure (SDI-Co), in which three principles for this new architecture are considered: open and transparent initiatives; a culture of participation, and the concept of inverse injection of spatial data. In this context, VGI data should not necessarily be seen as competition for authoritative data produced within an SDI initiative, but rather as an ally (Coetzee et al. 2013).

In the present work, from the proposed methodology, initial experiments were carried out for two municipalities (Valença and Petrópolis), both in the State of Rio de Janeiro, Brazil. The activities resulted in a semantic compatibility link between the ET-EDGV (DSG 2018b) – The Technical Specification for Vector Geospatial Data Structure, in Brazilian authoritative mapping – and OSM models, the identification of elements of the collaborative platform with the most significant potential to be incorporated into small scale authoritative mapping and the coverage and quality analyses of the possible data for integration. Thus, based on the results, this study aimed to describe challenges and present proposals for integrating data from collaborative mapping platforms with the authoritative mapping in Brazil.

It is essential to point out that, although data quality comparison confronting VGI data with authoritative databases is reasonably frequent in research, this paper brings a different approach. Here, the core focus is on finding gaps in the official data that VGI could fill, using case studies experiments to identify the main challenges and potentialities in the process. This perspective is important to create a methodology reproducible throughout the mapping of the country, considering the diversity of Brazilian authoritative mapping, that on a small scale, ranging from 1:25,000 to 1:250,000, is frequently outdated and available in absolutely heterogeneous environments (Silva and Camboim 2020).

There are several platforms or systems that provide volunteered geographic information in abundance. For this work, the OpenStreetMap (OSM) platform was chosen mainly for its degree of maturity regarding the structuring

of data, for its acceptance with the scientific community and for the whopping crowd of contributors in Brazil and worldwide. The OSM project, created in 2004, has more than 7 million registered users (October 2020) and around 1.1 million effective contributors who, on average, carry out more than 4.5 million changesets per day (OSMstats 2020), therefore keeping the database permanently updated and revised.

In addition to the millions of individual contributors, government institutions in about sixty countries (OSM 2020a), located on all continents, officially allow their data to be included on the OpenStreetMap platform. However, the OSM makes it clear that this does not imply that these data providers endorse, provide guarantees or accept any responsibility for the data contained in the platform.

Examples like these can be illustrated in bulk imports to OSM from official databases such as TIGER/Line in USA (U.S. Census Bureau 2002) and The Open Database of Buildings in Canada (Statistics Canada 2020). The TIGER data, produced by the US Census Bureau, includes many geographic features, especially the roads, which were imported into OSM in 2007 and 2008 and were important in populating, at that time, the nearly empty map of the United States (OSM 2020b). The Building Canada 2020 is a database of building footprints covering virtually all of Canada that contains about 11.8 millions polygon geometries in all Canadian ten provinces and three territories. This dataset is licensed by Microsoft in an open data license and was imported to OSM in 2018 and 2019 (Microsoft 2020). Despite its relevance and its recognized quality, for the purpose of the proposed methodology, this source of data is not under investigation, since in principle, they are already part of authoritative data.

More recently, another player has been highlighted in the context of OpenStreetMap: corporate editors, who are employees of companies and whose professional activity is editing data in OSM. Anderson, Sarkar and Palen (2019) show that in recent years the participation of corporate editors has grown dramatically, and part of this trend are some global giants such as Apple, Microsoft, Mapbox, Facebook and other large regional companies such as Grab (Singapore) and Telenav (United States).

The inclusion of government data, as well as the effective performance of large companies on the OpenStreetMap, represented a paradigm shift in the concept of OSM. This is due to the change in the initial understanding that the OSM was a database exclusively developed and maintained and by non-expert users. And more and more it is admitting a less amateur style, turning into a database with information from official sources and contributions, editions and validations carried out by people who are professionally dedicated to this task. These findings further reinforce the importance of studies aimed at using this type of content.

## 2. Methodology

### 2.1 Semantic compatibility between OSM and ET-EDGV

Semantic compatibility is the first step to make the use of this type of data feasible. It aims to make a semantic link between the features present in the OpenStreetMap system with one of the 225 object classes present in ET-EDGV (as detailed in the end of this section). So, in this process, the objective was to identify the individual characteristics of each type of OSM element, through the combination of “key” and “value”, and eventually the attributes in “tag”, and its correlation to the authoritative mapping data model described in ET-EDGV 3.0 version, as the examples shown in Table 1.

**Table 1:** Example of semantic compatibility between OSM and ET-EDGV

OpenStreetMap			ET-EDGV 3.0	
key	value	tag	Code	Class
highway	motorway	-	1.13.2	Trecho Rodoviário
highway	residential	-	2.2.27	Trecho Arruamento
boundary	administrative	admin_level = 8	1.4.11	Cidade
boundary	administrative	admin_level = 9	1.4.12	Distrito
power	tower	-	1.1.13	Torre Energia
natural	water	lake	1.3.12	Massa D'água

Semantic compatibility is an operation that identifies and documents the relationships between distinct data models. So, it ends up being a non-trivial task. An example of this is that the same combination of 'key + value' in the OSM can be related to different classes, so that the correspondence, in this case, needs additional information, the 'tag', to be defined. In addition, since data models are not necessarily concerned with the scale of representation, it is essential to keep in mind that elements represented on OSM may not find compatibility with the 1:25,000 and smaller scales that are the scope the national reference mapping in Brazil and of this study.


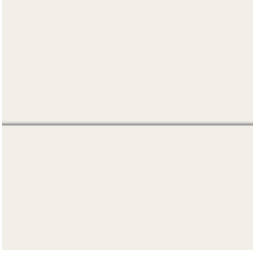
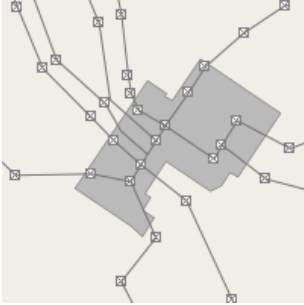
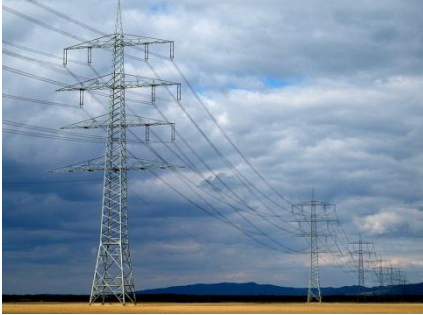
We present in Tables 2 and 3 some examples of information about geographic objects that allow understanding the semantic link between the data models.

**Table 2:** Information about 'Trecho\_Energia' class, in ET-EDGV 3.0

Code	Geometry	Description
1.1.15	line	Physical environment where the process of transmission and distribution of electricity is carried out.
Class	Figure	
Trecho_Energia	<p>The diagram shows a horizontal dashed line representing an electricity transmission line. On the left, a red dot is labeled 'Torre_Energia'. On the right, a red dot is labeled 'Complexo_Gerador_Energia_Eletrica'. A yellow box labeled 'Trecho_Energia' points to the line segment between the tower and the generator. Another yellow box labeled 'Trecho_Energia' points to the line segment extending to the right of the generator. The background is light green.</p>	

Source: Adapted and Translated to English, from DSG (2018a; 2018b).

**Table 3: Information about ‘power=line’ tag, in OpenStreetMap Wiki**

Tag	Element	Description
power=line	 way	High-voltage power lines used for power transmission, usually supported by towers or pylons.
Style	Mapnik	Figure
		

Source: Adapted from OSM (2020a).

Considering that the OSM is a map for global use, it is essential to highlight the difficulty in establishing a semantic structure and forms of classification and representation that suit the most diverse characteristics present on such a geographically, culturally, historically and economically heterogeneous planet.

The semantic compatibility between the models was a subject widely explored by Machado (2020). The author highlights that the conceptual model of the OSM is very well documented, illustrated and dynamic and emphasizes that the structure of the OSM is a contemporary example of the application of collective intelligence, given the information management and its platform mapping structure. However, as emphasized by Vandercastele and Devillers (2015), if on the one hand, the platform makes it possible to describe geographic objects in a rich way that meets local specificities, on the other hand, it creates semantic heterogeneities, since it allows a diversity of attributes or possible tags to describe the same object.

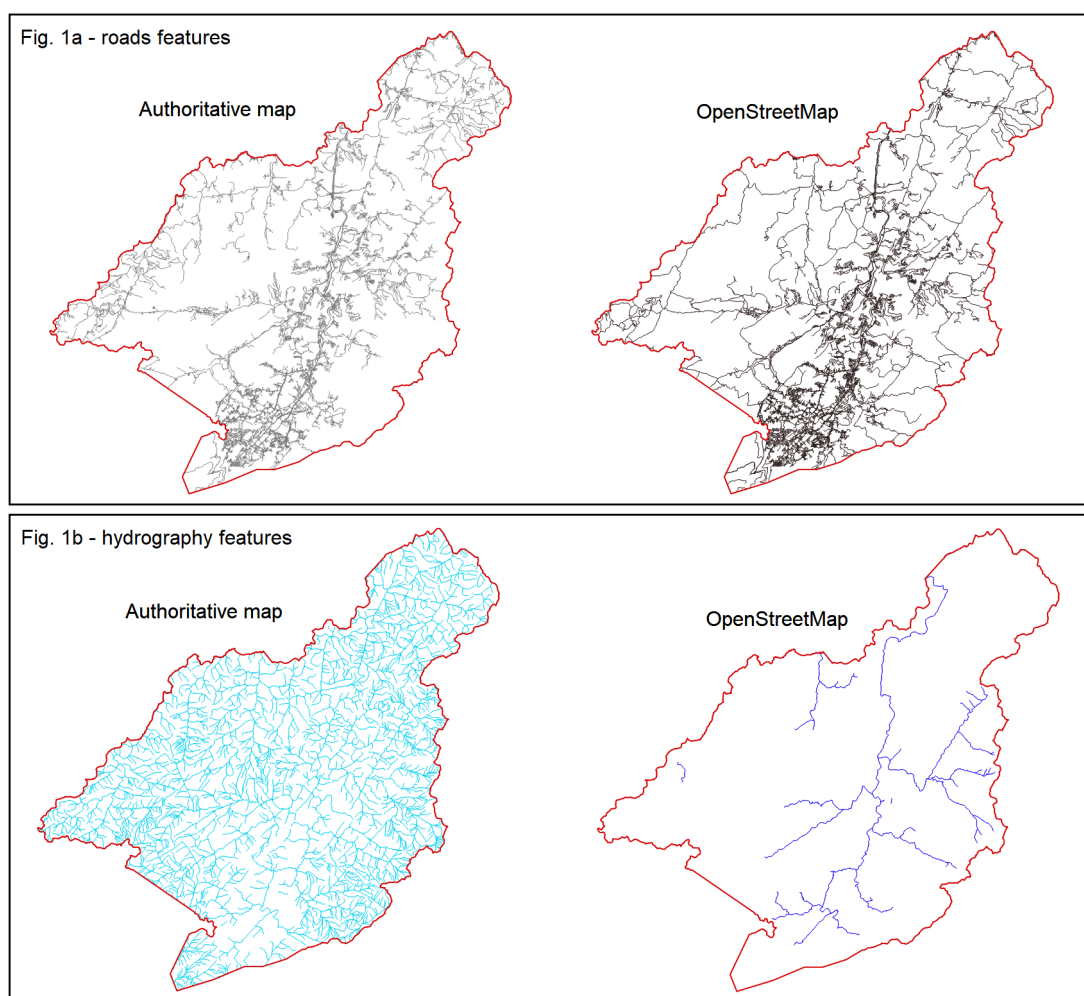
ET-EDGV, the Brazilian mapping data structure, based on OMT-G model, is defined in a more rigid and centralized way, approved by instances such as the former National Cartography Commission (CONCAR). Information producers mainly represented this committee, and the resulting data models are not very flexible. For example, there is a 9-year gap between versions 2 and 3 of ET-EDGV (2008 and 2017, respectively). ET-EDGV aims to standardize the official vectorial geospatial data structures of reference produced to compose cartographic bases related to the scales ranging from 1:1,000 to 1:250,000.

## 2.2. Selection of potential categories for integration

Considering that geospatial information from collaborative mapping platforms can be used to update authoritative mappings, a relevant aspect concerns the identification of which object classes present the most significant potential for integration.

The first approach is quantitative. The categories with more features added to the OSM can contribute more significantly to fill the gaps in the official mapping. It is important to highlight that the official base used in the experiments, covering the State of Rio de Janeiro, on scale 1:25,000, is considered one of the most updated in the country, made available in 2018.

Figure 1 shows a comparison between the features of the categories of information related to “roads” (Fig. 1a) and “hydrography” (Fig. 1b) for the municipality of Petrópolis, State of Rio de Janeiro. In the figure, the maps on the left represent objects of authoritative mapping and, on the right, elements of OpenStreetMap.



**Figure 1:** Comparison of the quantity of features in authoritative map and OSM.

From the illustrations in Figure 1, it can be seen that the category related to roads in the OSM has a large number of features represented (about 100% of quantitative equivalence with authoritative mapping, as quantitative presented in Table 7), while the category related to hydrography has few objects (about 9% of quantitative equivalence). The difference between these patterns in the OSM can be directly related to the way data is acquired and the use of the platform, for navigation purposes, for example.

This approach, more focused on determining the potential for integration concerning the representativeness of some categories, requires a preliminary analysis for each region of interest, since, due to the nature of the collaborative platform, the number of features present can be heterogeneous.

Another approach regarding the potential for integration is that made by Machado (2020) who proposes a method based on the semantic alignment between the ontologies of OSM and ET-EDGV 3.0, in which have tried to identify the key concepts and descriptions, values and labels of the collaborative platform and their correspondence in the structure of categories, classes, attributes and domain values of the authoritative mapping.

Through the cited study, among the eighteen categories of ET-EDGV, Machado (2020) identified results of high, medium, low or no percentages of semantic alignment, which give rise to some conclusions discussed below.



For categories composed of artificial features, located predominantly in urban areas, generally related to transportation, buildings, energy or communications, there is an abundance of data in the OSM, which results in high alignment between the models. Those categories related to natural features, which require technical knowledge or specific technologies to map, results in a medium or low alignment. And finally, there are classes whose definition is the responsibility of the government, for which there is low or no alignment.

The limitations related to the process of acquiring features by VGI, which may not have the quality requirements expected for the authoritative mapping, must be taken into account, in addition to the legal issues regarding the layers that are the responsibility of public sector, such as boundaries or Geodetic Control Points.

Taking into account the potential for integration based on expectation due to the high quantitative equivalence, reinforced by the analysis of the semantic alignment index, the classes 'Streets' (Trecho Arruamento) and 'Roads' (Trecho Rodoviário) of the authoritative mapping and the OpenStreetMap key 'highway' were chosen for the preliminary investigation of the proposed methodology.

## 2.3 Data manipulation and integration

Two study areas were defined to evaluate the proposed method, the municipalities of Valença and Petrópolis, which belong to the IBGE's continuous vector cartographic base (called in Portuguese Base Cartográfica Contínua) of the State of Rio de Janeiro, at 1:25,000 scale (IBGE 2018).

IBGE is the NMA co-responsible for authoritative mapping in Brazil and produced this authoritative cartographic base from photogrammetric coverage in 2006 and 2007 and updated by satellite images in 2018. This is the most up-to-date and largest-scale mapping available in Brazil.

In summary, the proposed methodology consists of the following steps.

- (a) Download elements of 'highway' key from OpenStreetMap;
- (b) Data acquisition of layers with 'line' geometry from the 'Streets' (Trecho Arruamento) and 'Roads' (Trecho Rodoviário) classes of the authoritative mapping;
- (c) Buffer generation, to serve as a positional uncertainty in comparison with the vectors from the OSM;
- (d) Comparison of the OSM vectors with the positional uncertainty (buffer) of the homologous lines in the authoritative mapping, to identify the coincidences and differences between them;
- (e) Separation of OSM vectors that are not present in the authoritative base and
- (f) Validation of the candidate features for integration.

Figure 2 highlights the two municipalities used and allows the identification of the study areas and the understanding of its representativeness in relation to the context of the entire country.

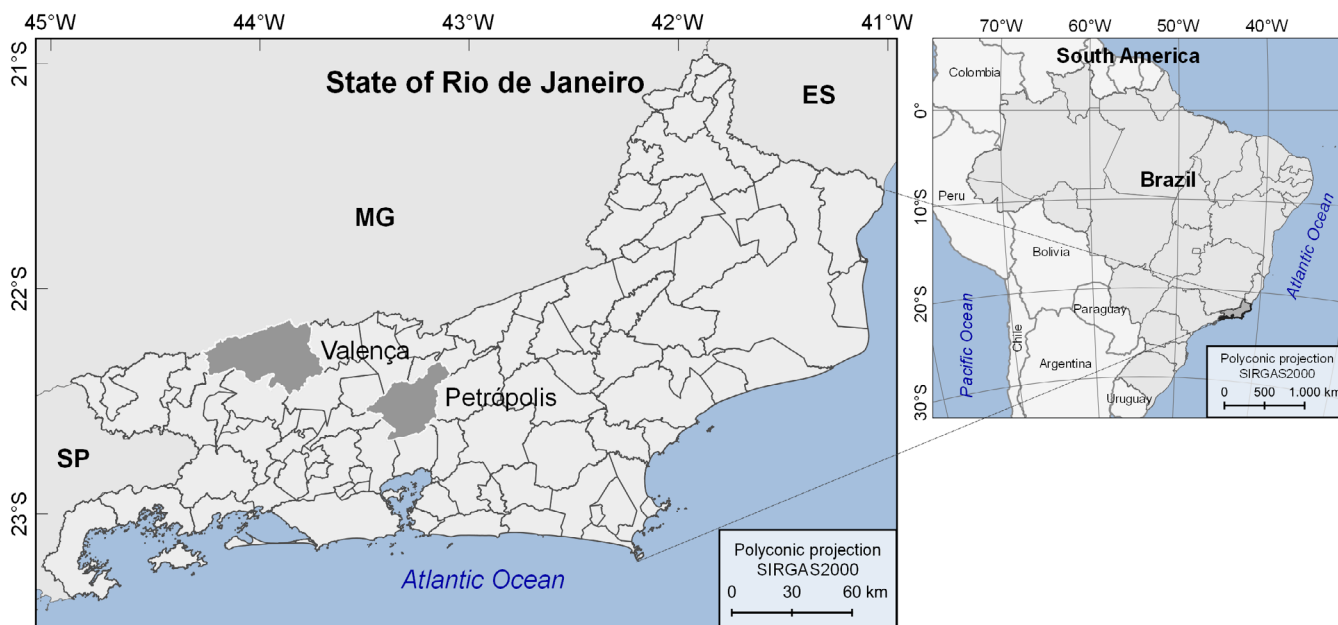


Figure 2: Study areas.

For the first area of interest (the municipality of Valença) the OpenStreetMap vectors that represent the features of the roads were acquired and identified on the platform by the key 'highway', as shown in Figure 3. The Coordinate Reference System is Universal Transverse Mercator (UTM), zone 23 (EPSG 31983). The total length of these vectors of the roads in that municipality is about 1,200 km.

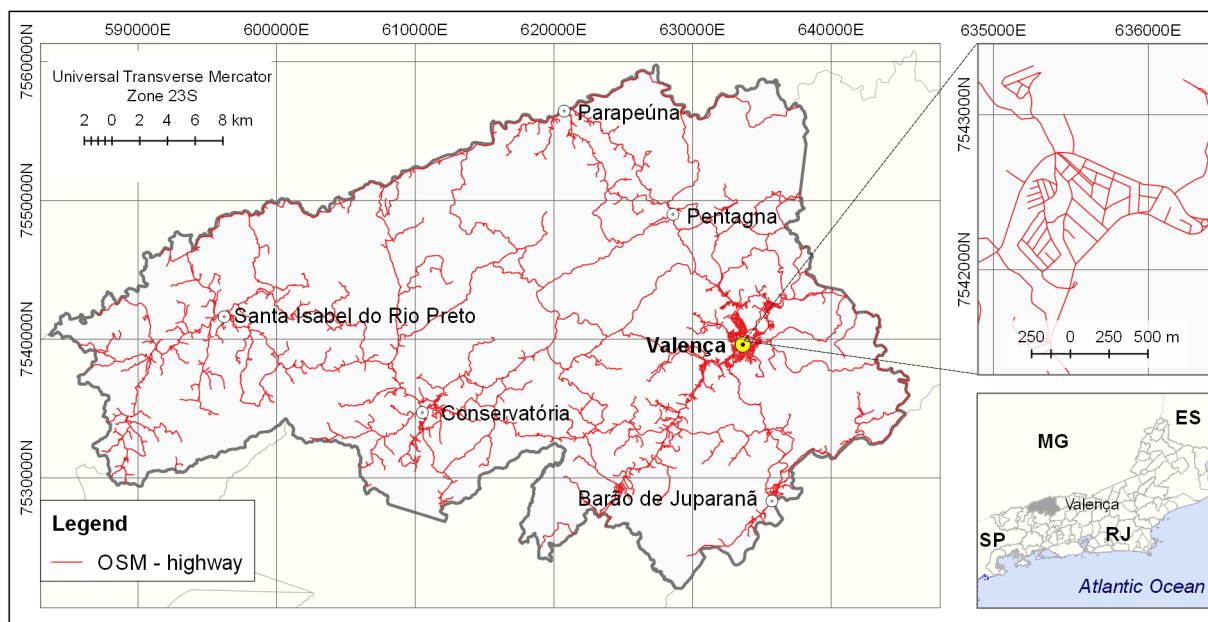
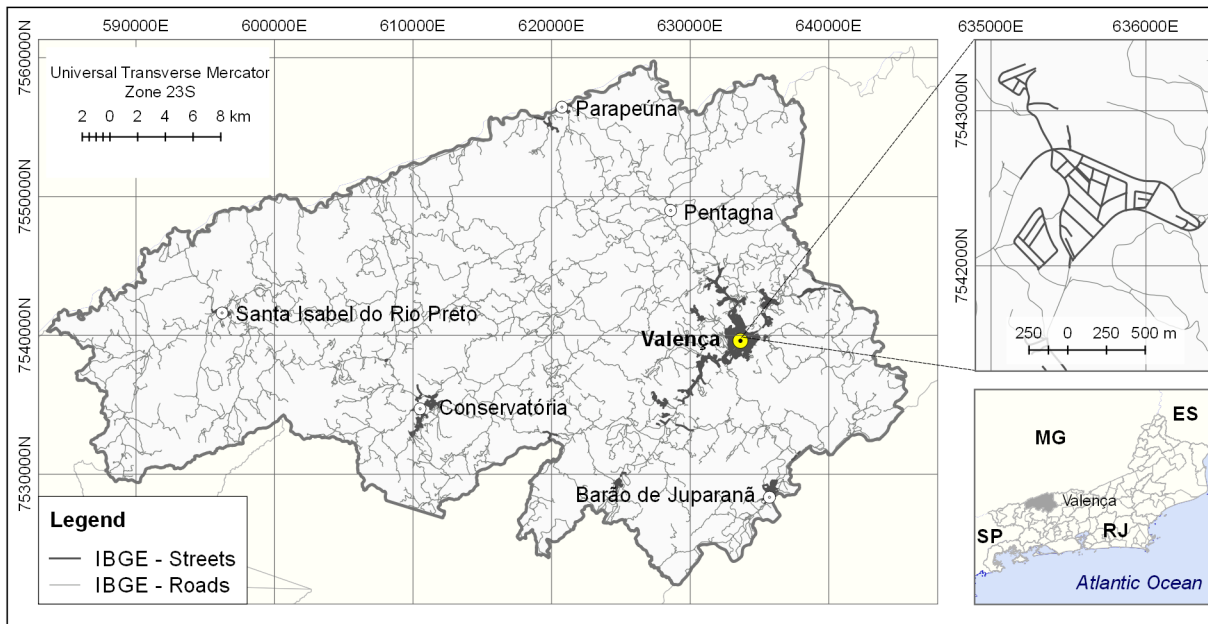


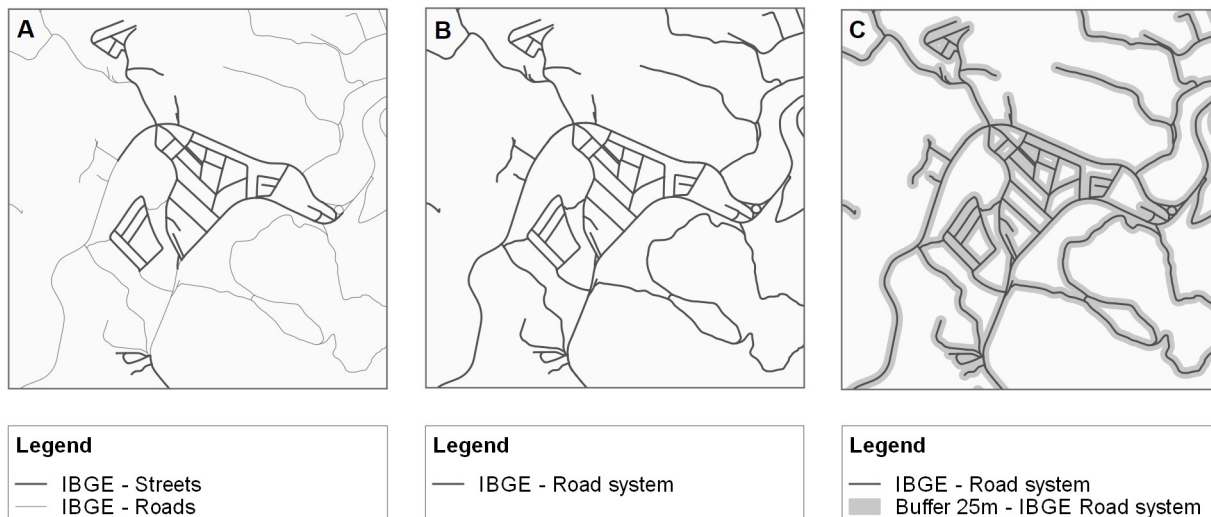
Figure 3: Features of the 'highway' from OpenStreetMap, for the municipality of Valença.

For the same study area, the vectors of the 'line' type geometry of the layers referring to the 'Streets' (Trecho Arruamento) and 'Roads' (Trecho Rodoviário) classes of continuous vector cartographic base of the State of Rio de Janeiro, in the 1:25,000 scale, were obtained as shown in Figure 4. The data were projected to Universal Transverse Mercator (UTM), zone 23 (EPSG 31983), to lead to the sequence of operations. The total length of these vectors of the roads in that municipality is about 2,040 km.



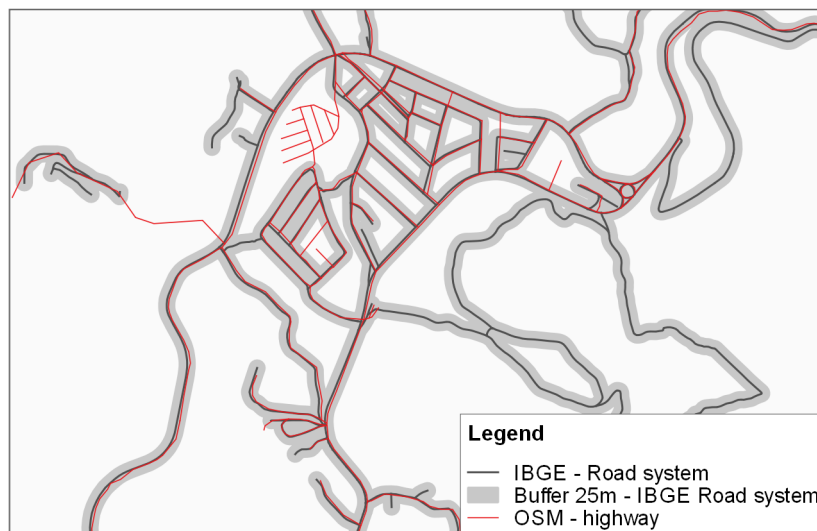
**Figure 4:** Features selected from 'Streets' and 'Roads' classes in the reference base, for the municipality of Valença.

The next step was the creation of a buffer of 1 mm on the mapping scale (corresponding to 25 meters) to serve as a range of uncertainty in comparison with the vectors coming from the OSM. This method is similar to the simple buffer method proposed by Mozas and Ariza (2011) and Santos (2015). The buffer size (25 m) was suggested based on findings of research by Haklay (2010) and Graser, Straub and Dragaschnig (2013) that found better results for 20 m buffers. Considering that those works were carried out to compare reference bases at larger scales, in this research we chose a slightly larger buffer for more flexibility. The sequence in the three progressive scenarios can be identified in Figure 5, respectively in windows A, B and C.



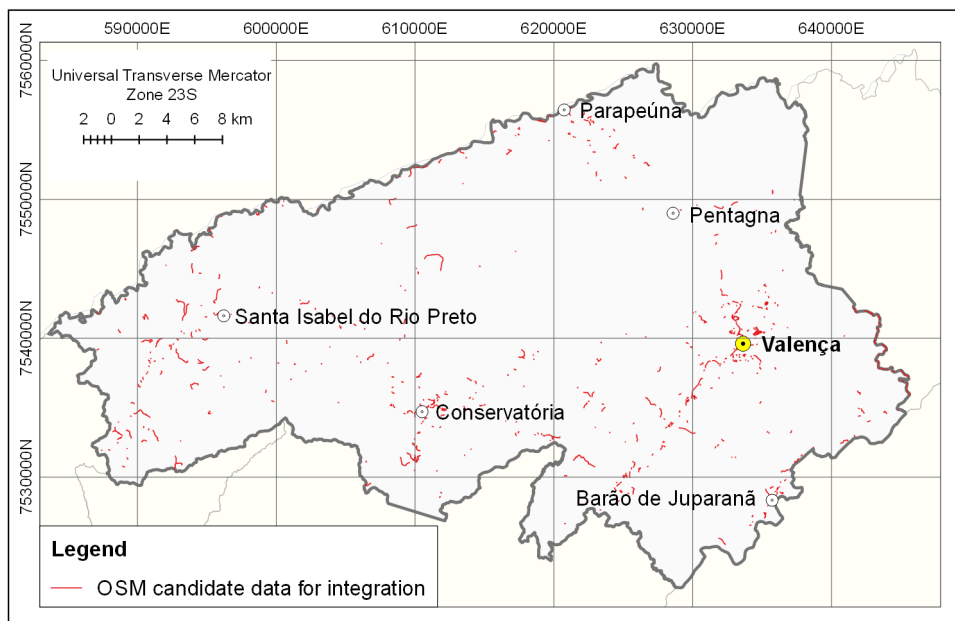
**Figure 5:** Separated layers (A), merged layers (B) and buffer of features in the authoritative map (C).

In the next step, vectors from OpenStreetMap were compared with vectors present in the authoritative base to identify the features that were present in both databases and those that were only present in the OSM. Figure 6 presents an example, in an enlarged view, that allows the identification and distinction between these two situations, that means, when the features of the OSM (represented in red) coincide with the buffer (represented in gray) or when there is no correlation with the authoritative mapping.



**Figure 6:** Correlations and differences between the authoritative map and OpenStreetMap.

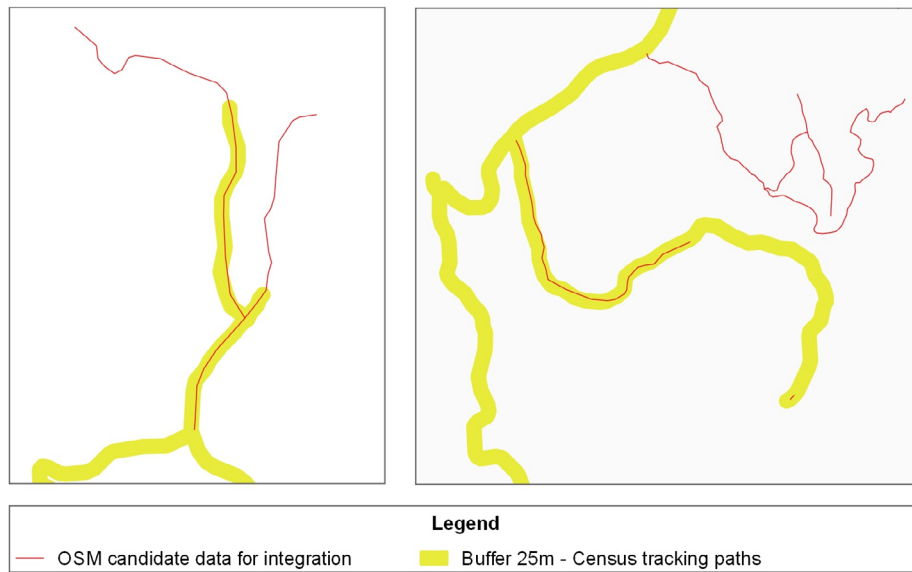
In the next step, we selected the features from OpenStreetMap not represented in the authoritative database (vectors in red, not coincident with the buffer in gray), therefore classified as candidate data for integration. Figure 7 shows all vectors of the OSM that are not present in the authoritative base in the area of the municipality of Valença. These vectors have a total of about 94 km in length.



**Figure 7:** OSM highway vectors not present in the authoritative base.

The next phase was the validation of the candidate vectors for integration. For this task, we used the vectors of the tracking paths of IBGE’s workforce in the 2017 Agricultural Census. In this project, the workers recorded in navigation GNSS receivers all the routes that they passed through the activities of the Census in 2017. Therefore, a large updated and reliable dataset of roads was obtained, which made it possible for comparison and validation of VGI. In this step, a buffer of the distance equivalent to 1 mm on the mapping scale (corresponding to 25 meters) was also considered for the intersection of homologous features. Figure 8 exemplifies the validation process carried out, in which, part of the elements of OSM coincides with the buffer of IBGE Census tracking paths, and another part does not. The sum of the length of the coincident vectors is about 49 km that represents 52% of tested roads.

Thus, by the proposed methodology, we can observe that the coincident elements have consistent quality and may be integrated with the authoritative mapping.



**Figure 8:** Validation process of OSM vectors based on comparison with Census tracking paths.

The same proposed methodology was carried out similarly for the municipality of Petrópolis, and the results for both study areas are presented in the following section.

### 3. Results and Discussions

As a result of the first part of the evaluation, Table 4 presents the quantitative length of roads vectors on the authoritative mapping (IBGE) and the collaborative platform (OSM), the differences between them and what percentage of these data has potential for acquisition after basic validation.

**Table 4:** Quantitative and percentage of features for Valença and Petrópolis

Municipality of Valença			
Element	Length (km)	Percentage (%)	Compared to
Road system (Authoritative map IBGE)	2.042,092	-	-
Highway (OSM)	1.200,163	58,8	IBGE Road system
OSM highway no present in IBGE	94,344	4,6	IBGE Road system
Candidate data of OSM validated by Census tracking paths	49,463	52,4	OSM highway no present in IBGE
Municipality of Petrópolis			
Element	Length (km)	Percentage (%)	Compared to
Road system (Authoritative map IBGE)	1.901,244	-	-
Highway (OSM)	1.898,830	99,9	IBGE Road system
OSM highway no present in IBGE	254,280	13,4	IBGE Road system
Candidate data of OSM validated by Census tracking paths	42,195	16,6	OSM highway no present in IBGE

Due to the quantities presented in Table 4, it was possible to verify that the representativeness of the elements of the roads of the OpenStreetMap in relation to the analogous content in the authoritative mapping, reaches almost 59% of the quantity in length for the municipality of Valença. Among the elements of the OSM, about 94 km of highway are not present in the authoritative base, which could mean an increase of 4.6% of this type of feature. For the municipality of Petrópolis, that the representativeness of the elements of the roads in the OpenStreetMap in relation to the analogous content in the reference base reaches almost 100% of the quantity in extension. Among the elements of the OSM, about 254 km are not present in the authoritative base, which could mean an increase of 13.4% of this type of feature.

As the municipality of Valença has a small urban area, there was a massive difference in the proportion of vectors that can be acquired in the rural area (95.4%) in relation to the urban area (4.6%). Furthermore, it is important to consider that most of the tracking paths of census took place in rural areas, because it was the region most visited by interviewers in the census. There is a significant difference when this analysis, according to Tables 5, distinguishes the elements of the roads belonging to urban and rural areas. The validation with this kind of data would be possible for about 18% of urban roads and 54% of rural roads. This analysis shows that more than 52% (Table 4) of those elements candidates for acquisition were validated by the tracking paths resulting from the census activities.

Since the municipality of Petrópolis has a large urban area, there was a much smaller proportion difference between the vectors that can be acquired in the rural area (63.6%) compared to the urban area (36.4%). Nevertheless, according to Table 5, the elements of the roads belonging to urban areas and subject to validation with this resource would be about 8% of urban roads and 21% of rural roads. In contrast to the evaluation for Valença, a much less significant quantity (16.6%- Table 4) of elements that were candidates for the acquisition was obtained, which could be validated by the tracking paths resulting from census activities.

**Table 5:** Quantitative and percentage of OSM highway no present in IBGE road system

<b>Valença - Rural area</b>			
<b>Element</b>	<b>Length (km)</b>	<b>Percentage (%)</b>	<b>Compared to</b>
Rural OSM vectors no present in IBGE	90,062	95,4	OSM highway no present in IBGE (94,344 km)
Candidate data of OSM validated by Census tracking paths	48,681	54,0	Rural OSM vectors no present in IBGE
<b>Valença – Urban area</b>			
<b>Element</b>	<b>Length (km)</b>	<b>Percentage (%)</b>	<b>Compared to</b>
Urban OSM vectors no present in IBGE	4,282	4,6	OSM highway no present in IBGE (94,344 km)
Candidate data of OSM validated by Census tracking paths	0,782	18,3	Urban OSM vectors no present in IBGE
<b>Petrópolis – Rural area</b>			
<b>Element</b>	<b>Length (km)</b>	<b>Percentage (%)</b>	<b>Compared to</b>
Rural OSM vectors no present in IBGE	161,792	63,6	OSM highway no present in IBGE (254,280 km)
Candidate data of OSM validated by Census tracking paths	34,489	21,3	Rural OSM vectors no present in IBGE
<b>Petrópolis – Urban area</b>			
<b>Element</b>	<b>Length (km)</b>	<b>Percentage (%)</b>	<b>Compared to</b>
Urban OSM vectors no present in IBGE	92,488	36,4	OSM highway no present in IBGE (254,280 km)
Candidate data of OSM validated by Census tracking paths	7,706	8,3	Urban OSM vectors no present in IBGE

The availability of data from the roads on OpenStreetMap showed off is quite significant, with examples of almost 60% (Valença) and 100% (Petrópolis) of the amount of length of roads present in the authoritative base. As the analysis was focused on OSM data that are not present in the authoritative database, potential increases of 4.6% and 13.4% in the length of roads in this category were calculated for the municipalities of Valença and Petrópolis, respectively. The difference found may be related to the hypothesis that regions with more population and more tourism activities should receive more contributions to collaborative platforms, endorsing the findings of previous studies as Neis, Zielstra and Zipf (2013). Concerning the validation process, it was identified, as expected, that, due to the existence of more data on the tracking paths of the census in rural regions, the validation of the roads from this dataset was more successful in municipalities with a higher proportion of rural areas, as in the case of Valença. However, the legal characterization of rural and urban spaces in Brazil still has many divergences in its understanding and in the definition itself.

In many cases, the most common data sources for validation processes can be large geospatial datasets generated for specific projects or activities (such as the Agricultural Census or school transport network), cartographic data at local or state administrative level, or features identification in orthophotos or high resolution satellite images. Another possibility to be considered is the validation from data confirmation by expert and qualified users with a high reputation in collaborative mapping projects

Analyzes and considerations presented here were based on a sample of tests. Given the various possibilities surrounding the subject, they address us to discussions and challenges that are presented below.

Although the evaluations were carried out for the extension of only two municipalities and compared with the dataset of only one state (Rio de Janeiro) that represents about 0.5% of the country, we consider that the potential that these results may be even higher in the other regions of the country. This is due to the fact that in this case, the reference database is a recent mapping, on the scale of 1:25,000 and updated in the year 2018. When compared with the reality of the country, it is observed that the availability of such an updated and detailed dataset is an exception, according to Silva and Camboim (2020) that calculated that only 5.5% of Brazil's territory is mapped on the 1:25,000 scale, and in this scale, about 68% were less than 12 years old.

Given this reality, there is an excellent opportunity to extend the specific findings of this research to a much larger area and quite diverse scenarios, since most of the country is mapped on smaller scales (1:50,000 and 1:100,000) and, mostly (more than 60%) at least 20 years of age. Additionally, the situation of other developing countries is in Brazil's economy, and these reflections can help to think about integration in a broader global context than previous research.

The findings of this work can be synthesized in a SWOT analysis in order to classify strengths, weaknesses, opportunities and threats, as well as to correlate the challenges and proposals that resulted from this methodology, according to Table 6.

**Table 6:** SWOT analysis

<b>Strengths</b>	<b>Weakness</b>
Promote the inclusion of local knowledge and public participation	Some classes of objects have few features
Assist in data generation in places where there is a lack of information	Some classes of objects have limitations for VGI acquisition
Make it easier to update mapping with higher frequency and speed	Validation depends on the availability of data for this purpose or potentially costly processes
Use the development of artificial intelligence and other tools available for collaborative mapping	Need to individualize feature metadata and maintain real-time automated integration between databases
<b>Opportunities</b>	<b>Threats</b>
Can indicate regions in more urgent need for mapping	Disinterest in integrating VGI data or using it in SDI
May encourage the development of local SDIs and active communities in VGI	Traditionally marginalized or low-data areas in official maps also often have few contributors to
Alternative means for validating VGI data may arise	VGI may compromise SDI frameworks by not respecting interoperability standards and limitations
It may favour integration with other governmental entities	The transfer of responsibility of mapping to society can lead to unavailability of resources for this purpose even worse by the government
Potential positive synergies between the OSM community and official data producers	Little knowledge about users' motivations and little action in fostering expansion and improvement of contributions

Performing a complex and as challenging task as integrating VGI and authoritative data, from bases with so many heterogeneities, in a country with continental dimensions needs a lot of planning. In this way, the SWOT analysis represents a tool to address future research and assist strategic planning for governmental actions that may guide future steps.

This SWOT analysis was developed based on the Brazilian reality, however, it is quite possible that it is also suitable for some nations in a similar situation, especially those developing countries.

## 4. Conclusion

From the analysis of the promising experiments carried out, it is possible to affirm that the proposed methodology is feasible for its purpose of complementing and updating the reference bases in the country from easily accessible and low-cost content. As collaborative systems do not have a uniform number of contributions it is understood that for each region, the potential for increment may vary for each class as well as for each mapping scale.

As challenges for expanding the present methodology, further studies are needed to assess the quality parameters of the data selected for integration, to attest their compatibility with the quality of the authoritative mapping. Another challenge regards the definition of cartographic generalization processes since the features from the collaborative platforms are acquired in a variety of ways, with different precision and, therefore, may need to be adapted to be represented on a defined scale.

As proposals for data integration studies, it is suggested that the research be extended to the other classes belonging to the authoritative mapping, and the potential for integration for each of them should be evaluated in detail, according to the criteria presented in section 2.2.



Additionally, because of the diversity of object classes, as well as the intention to proceed with data integration for the whole country, it is suggested to define protocols for the validation of the acquirable data, from different types and different data sources.

In a global view, SDIs initially came to bring the idea of standardization, openness and transparency to the creation and access to geospatial data, mainly those produced by the government. Over time, technology has expanded this vision to encompass all citizens as potential mappers and holders of the knowledge of their local reality. The challenge now is to create methods so that synergy between these communities can result in an understanding of the territory in a more agile, accurate and consistent. Thus, developing countries such as Brazil can have at their disposal spatial information that is so necessary for a complex, dynamic territorial management reality and with a series of social and environmental issues to address.

## ACKNOWLEDGMENT

This article results from the initial part of the doctoral research carried out by Leonardo Scharth Loureiro Silva under the supervision by Silvana Philippi Camboim, which culminated in February 2020 in the defense of research qualification for the future thesis.

We thank Thais Christ for the linguistic review on an earlier version of the paper.

## AUTHOR'S CONTRIBUTION

Leonardo Scharth Loureiro Silva wrote the manuscript and carried out the experiments. Silvana Philippi Camboim contributed to the methodological structuring and to the revision of the final version.

## REFERENCES

- Al-Bakri, M. and Fairbairn, D. 2012. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. *International Journal of Geographical Information Science*, 26(8), pp.1437–56.
- Albuquerque, J. P. et al. 2015. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science*, 29(4), pp.667–89.
- Anderson, J. Sarkar, D. and Palen, L. 2019. Corporate Editors in the Evolving Landscape of OpenStreetMap: A Close Investigation of the Impact to the Map & Community. In: *Proceedings of the Academic Track at the State of the Map 2019*. Heidelberg, Germany, 21-23 Sep 2019.
- Basiouka, S. Potsiou, C. and Bakogiannis, E. 2015. OpenStreetMap for cadastral purposes: an application using VGI for official processes in urban areas. *Survey Review*, 47(344), pp.333-41.
- Borba, R. L. R. et al. 2015. Uma proposta para a nova geração de infraestrutura de dados espaciais. *Revista Brasileira de Cartografia*, 67(6), pp.1145-66.
- Bordogna, G. et al. 2016. Semantic Interoperability of Volunteered Geographic Information based on Contextual Knowledge. In: *AGILE 2016*. Helsinki, Finland, 14-17 Jun 2016.

- Brovelli, M. A. et al. 2016. Positional accuracy assessment of the OpenStreetMap buildings layer through automatic homologous pairs detection: the method and a case study. In: *XXIII ISPRS Congress*. Prague, Czech Republic, 12-19 Jul 2016.
- Brovelli, M. A. et al. 2017. Towards an Automated Comparison of OpenStreetMap with Authoritative Road Dataset. *Transactions in GIS*, 21(2), pp.191-206.
- Coetzee, S. et al. 2013. Sharing and Coordinating SDI in the Age of Crowdsourcing and Mobile Technologies. In: *26th International Cartographic Conference*. Dresden, Germany, 25-30 Aug 2013.
- Coleman, D. J. 2013. Potential Contributions and Challenges of VGI for Conventional Topographic Base-Mapping Programs. In: D. Z. Sui, S. Elwood and M. F. Goodchild, eds. 2013. *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. Springer Netherlands. pp.245-263.
- Dorn, H. Törnros, T. and Zipf, A. 2015. Quality Evaluation of VGI Using Authoritative Data – A comparison with land use data in Southern Germany. *ISPRS International Journal of Geoinformation*, 4(3), pp.1657-71.
- DSG – Diretoria do Serviço Geográfico. 2018a. *ET-ADGV 3.0 – Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais*. [online] Available at: <<http://www.geoportal.eb.mil.br/portal/inde2?id=140>> [Accessed 25 October 2019].
- DSG – Diretoria do Serviço Geográfico. 2018b. *ET-EDGV 3.0 – Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais*. [online] Available at: <<http://www.geoportal.eb.mil.br/portal/inde2?id=139>> [Accessed 25 October 2019].
- Fonte, C. C. et al. 2018. Classification of Building Function using available sources of VGI. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Delft, The Netherlands, 1-5 Oct 2018.
- Forghani, M. and Delavar, M. R. 2014. A Quality Study of the OpenStreetMap Dataset for Tehran. *ISPRS International Journal of Geoinformation*, 3(2), pp.750-63.
- Girres, J.-F. and Touya, G. 2010. Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, 14(4), pp.435–59.
- Goodchild, M. F. 2007. Citizens as sensors: the world of volunteered geography. *Geojournal*, 69(4), pp.211-21.
- Graser, A. Straub, M. and Dragaschnig, M. 2013. Towards an Open Source Analysis Toolbox for Street Network Comparison: Indicators, Tools and Results of a Comparison of OSM and the Official Austrian Reference Graph. *Transactions in GIS*, 18(4), pp.510-26.
- 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), pp.682–703.
- Harding, J. et al. 2009. Usable geographic information – what does it mean to users? In: *Proceedings of the AGI GeoCommunity'09*. Stratford-upon-avon, United Kingdom, 22-24 Sep 2019.
- Harvey, F. and Tulloch, D. 2006. Local-government data sharing: Evaluating the foundations of spatial data infrastructures. *International Journal of Geographical Information Science*, 20(7), pp.743-68.
- Hecht, R. Kunze, C. and Hahmann, S. 2013. Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. *ISPRS International Journal of Geoinformation*, 2013(2), pp.1066-91.
- IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. 2018. *IBGE launches 2018 version of continuous vector cartographic base of Rio de Janeiro*. [online] Available at: <<https://agenciadenoticias.ibge.gov.br/en/agencia-press-room/2185-news-agency/releases-en /21578-ibge-lanca-a-versao-2018-da-base-cartografica-vetorial-continua-do-estado-do-rio-de-janeiro-2>> [Accessed 15 October 2020].
- Johnson, P. A. 2017. Models of direct editing of government spatial data: challenges and constraints to the acceptance of contributed data. *Cartography and Geographic Information Science*, 44(2), pp.128-38.
- Korris, E. Niknami, L. and McCartney, E. 2017. Assessment of Volunteered Geographic Information Data Quality in The National Map Corps Project of the US Geological Survey (USGS). In: M. Leitner, J. J. Arsanjani, eds. 2017. *Citizen Empowered Mapping*. Geotechnologies and the Environment, vol 18. Springer, Cham.

- Machado, A. A. 2020. *Compatibilização semântica entre o modelo de dados do OpenStreetMap e a Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV)*. PhD thesis. Universidade Federal do Paraná.
- Massa, P. and Campagna, M. 2016. Integrating Authoritative and Volunteered Geographic Information for spatial planning. In: C. Capineri, M. Haklay, H. Huang, V. Antoniou, J. Kettunen, F. Oostermann, R. Purves, eds. 2016. *European Handbook of Crowdsourced Geographic Information*. Ubiquity Press, London, pp.401-18.
- Microsoft. 2020. *Canadian Building Footprints*. [online] Available at: <<https://github.com/Microsoft/CanadianBuildingFootprints>> [Accessed 3 Nov 2020].
- Mooney, P. and Corcoran, P. 2011. Can Volunteered Geographic Information Be a Participant in eEnvironment and SDI? In: J. Hřebíček, G. Schimak and R. Denzer, eds. 2011. *Environmental Software Systems. Frameworks of eEnvironment*, vol 359. Springer, Heidelberg.
- Mozas, A. T. and Ariza, F. J. 2011. New method for positional quality control in cartography based on lines. A comparative study of methodologies. *International Journal of Geographical Information Science*, 25(10), pp.1681-95.
- Neis, P. Zielstra, D. and Zipf, A. 2013. Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet*, 5, pp.282-300.
- Olteanu-Raimond, A.-M. et al. 2017a. The Scale of VGI in map production: a perspective on European National Mapping Agencies. *Transactions in GIS*, 21(1), pp.74-90.
- Olteanu-Raimond, A.-M. et al. 2017b. VGI in National Mapping Agencies: Experiences and Recommendations. In: G. Foody, L. See, S. Fritz, P. Mooney, A.-M. Olteanu-Raimond, C. C. Fonte, V. Antoniou, eds. 2017. *Mapping and the Citizen Sensor*. London: Ubiquity Press, 2017. pp.299-326.
- OSM – OpenStreetMap, 2020a. *Wiki, Contributors*. [online] Available at: <<https://wiki.openstreetmap.org/wiki/Contributors#Others>> [Accessed 17 May 2020].
- OSM – OpenStreetMap, 2020b. *Wiki, TIGER*. [online] Available at: <<https://wiki.openstreetmap.org/wiki/TIGER>> [Accessed 3 Nov 2020].
- OSMstats, 2020. *Statistics of the free wiki world map*. [online] Available at: <<https://osmstats.neis-one.org/>> [Accessed 10 Oct 2020].
- Santos, A. P. 2015. *Controle de qualidade cartográfica: Metodologias para avaliação da acurácia posicional em dados espaciais*. PhD thesis. Universidade Federal de Viçosa.
- Sester, M. et al. 2014. Integrating and Generalising Volunteered Geographic Information. In: D. Burghardt, C. Duchêne, W. Mackaness, eds. 2014. *Abstracting Geographic Information in a Data Rich World: Methodologies and Applications of Map Generalisation*. Cham: Springer International Publishing, 2017. pp.119-55.
- Silva, L. S. L. and Camboim, S. P. 2020. Brazilian NSDI ten years later: Current overview, new challenges and propositions for national topographic mapping. *Bulletin of Geodetic Sciences*, 26(4): e2020018.
- Statistics Canada. 2020. *The Open Database of Buildings*. [online] Available at: <<https://www.statcan.gc.ca/eng/lode/databases/odb>> [Accessed 3 November 2020].
- U.S. Census Bureau. 2002. *The UA Census 2000 TIGER/Line Files Technical Documentation*. Washington, DC.
- USGS – UNITED STATES GEOLOGICAL SURVEY. 2019. *Volunteers collect and edit over 400K structure points*. [online] Available at: <<https://www.usgs.gov/news/volunteers-collect-and-edit-over-400k-structure-points>> [Accessed 9 October 2020].
- Vandecasteele, A. and Devillers, R. 2015. Improving Volunteered Geographic Information Quality Using a Tag Recommender System: The Case of OpenStreetMap. In: J. J. Arsanjani, A. Zipf, P. Mooney, M. Helbich, eds. 2015. *OpenStreetMap in GIScience: Experiences, Research and Applications*. Lecture Notes in Geoinformation and Cartography. Springer, Cham, pp.59-80.

## Erratum

**On all pages where it reads:**

*Boletim de Ciências Geodésicas, 27(1)*

**Should read:**

*Boletim de Ciências Geodésicas, 27(spe)*

Bulletin of Geodetic Sciences, 27(spe): e2021007, 2021