Revista Brasileira de Cartografia
ISSN 1808-0936 | https://doi.org/10.14393/revbrascartogr
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto

Semantic Alignment of Official and Collaborative Geospatial Data: A Case Study in Brazil

Alinhamento Semântico entre Dados Geoespaciais Oficiais e Colaborativos: Um estudo de caso no Brasil

Adriana Alexandria Machado[1], Silvana Philippi Camboim[2]

1 Universidade Federal do Paraná, Curitiba, Brasil. adri.alexandria@gmail.com
ORCID: https://orcid.org/0000-0003-1013-2050
2 Universidade Federal do Paraná, Curitiba, Brasil. silvanacamboim@gmail.com
ORCID: https://orcid.org/0000-0003-3557-5341

Recebido: 01.2024 | Aceito: 04.2024

Abstract: Geospatial data is crucial for sustainable development, but obtaining up-to-date and high-quality data is challenging in many regions, including Brazil. Collaborative mapping on platforms such as OpenStreetMap (OSM) has produced updated and open geospatial data, especially in urban areas, but its quality is heterogeneous. In addition, semantic interoperability is challenging when integrating OSM data with authoritative geospatial data. This article presents a procedure for semantic alignment between two conceptual models within a conflation process to elicit background knowledge for geospatial data integration. The first model is the Technical Specification for Structuring Vector Geospatial Data (ET-EDGV 3.0) in Brazilian Portuguese, and the second is the OSM model with tags mainly in English. The alignment produced a table combining the ET-EDGV classes, attributes, domains, and geometries with the OSM tags and elements. The semantic alignment was tested in two study areas to check the thematic accuracy of transportation data imported from OSM compared to the data in the reference database. The study found that the best percentage of segments correctly classified by alignment was for "highway=trunk" tags (98.27%) and "highway=primary" (98.20%), corresponding to road and highway segments, and for the "highway=residential" tag (76.20%), corresponding to sections of residential streets. The study also identified factors that may contribute to low accuracy rates, including ambiguous semantic descriptions and the need for local context analysis. This research contributes to adding collaborative data to the official mapping, a relevant alternative for updating and supplementing reference mapping that can be applied in other geographical contexts.
Keywords: Semantic interoperability. Semantic alignment. Integration between geospatial databases. Data conflation.

Resumo: Os dados geoespaciais são cruciais para o desenvolvimento sustentável, entretanto obter dados atualizados com alta qualidade é um desafio em muitas regiões, incluindo o Brasil. O mapeamento colaborativo de plataformas comoo OpenStreetMap (OSM) produz dados geoespaciais atualizados e abertos, especialmente em áreas urbanas, entretanto, sua qualidade é heterogênea. A interoperabilidade semântica é outro desafio para integrar dados OSM à dados oficiais. Este artigo apresenta um procedimento de alinhamento semântico entre dois modelos conceituais dentro de um processo de fusãopara obter conhecimento prévio para integração de dados. O primeiro modelo é a Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV 3.0) e o segundo é o modelo do OSM com *tags* em inglês. O alinhamento produziu uma tabela combinando as classes, atributos, domínios e geometrias da ET-EDGV com as *tags* e elementos do OSM. Este alinhamento foi testado em duas áreas de estudo para verificar a precisão temática dos dados de transporte importados do OSM em comparação com os dados oficiais. O estudo constatou que o melhor percentual de dados classificados corretamente pelo alinhamento foi para as *tags* "highway=trunk" (98,27%) e "highway=primary" (98,20%), correspondentes aos trechos rodoviários e rodovias, e"highway=residential"(76,20%) correspondente a trechos de vias residenciais. A pesquisa identificou ainda, fatores que podem contribuir para baixas taxas de precisão, como descrições semânticas ambíguas e necessidade de análise do contexto local. Esta pesquisa contribui para a agregação de dados colaborativos ao mapeamento oficial, para atualização e completude, uma alternativa que pode ser aplicada em outros contextos geográficos.
Palavras-chave: Interoperabilidade semântica. Alinhamento semântico. Integração entre bancos de dados geoespaciais. Fusão de dados.

1 INTRODUCTION

Due to the low investment in cartography in Brazil, the authoritative geospatial data covers the territory mainly at the small and medium scales and is also often outdated. Moreover, large-scale reference mapping, necessary for urban planning, is still scarce in the country (CAMBOIM et al., 2015). On the other hand, collaborative mapping of platforms such as OpenStreetMap (OSM) produces updated and open geospatial data, especially in urban areas. Although its quality is heterogeneous, studies have shown that it is satisfactory for many usessuch as change detection, collection of new data, photointerpretation, web-based tools,vernacular names, toponymy, use of metadata and others, what is already happening in several countries like Germany, Belgium, Canada, Great Britain and the United States of America (DORN et al., 2015; ELIAS and FERNANDES, 2021; SALVUCCI and SALVATI, 2022; ZHANG and MALCZEWSKI, 2017). Therefore, OSM's collaborative data can be a relevant alternative for updating and supplementing reference mapping. However, achieving interoperability between data and geospatial information systems is not a trivial task due to the complexity of geospatial information and the variety of modelling possibilities.

At the semantic level, the lack of common conceptual models implies problems in data exchange between organisations, including data distortion, poor information quality, loss in attribute definitions, and even georeferencing (LIMA et al., 2002; WACHE et al., 2001). Thus, establishing a conceptual model provides a path to overcome the challenge of semantic interoperability. Furthermore, a conceptual model provides a coherent and systematic description of the content and organisation of a dataset. However, when working with distinct conceptual models, semantic compatibility between them is required to integrate multiple databases. For this reason, performing a semantic compatibility match (SIEBER and JOHNSON, 2013; NOVACK et al., 2019; TRAJCEVSKI et al., 2020) among their respective definitions is necessary.

However, formalising and associating concepts with a shared meaning makes this mission challenging. Moreover, there may be variations in the concepts depending on geographical context or usage, as definitions result from human perception and cognition, social consensus and state of knowledge (NOVACK et al., 2019).

This article presents a procedure for semantic alignment (TRAJCEVSKI et al., 2020; XAVIER et al., 2017; ZHANG et al., 2010) between two conceptual models within a conflation process. Semantic alignment is used to elicit background knowledge for geospatial data integration. In this research, we propose to apply it in the context of geospatial authoritative and collaborative data interoperability and as a first step towards a subsequent formalisation of an ontology.

A conflation process integrates the best-quality components of different data sources to build a better-combined dataset or provide complementary information that cannot be obtained from a single data source (VILCHES-BLÁZQUEZ and RAMOS, 2021). The first model chosen is the one established by the National Cartography Commission for Brazilian topographic mapping, the Technical Specification for Structuring Vector Geospatial Data (ET-EDGV 3.0) (CONCAR, 2018) in Brazilian Portuguese. The second is the model proposed and maintained by the OSM community (OpenStreetMap Wiki, 2023) with tags mainly in English. This process produced a table combining the ET_EDGV classes, attributes, domains and geometries with the OSM tags and elements. Documenting the definitions, their respective sources, and the the alignment graph, we tested the alignment by applying it to two study areas to check the thematic accuracy of the data imported from OSM compared to the data in the reference (official) database. These procedures may provide a basis for further integrating heterogeneous data sources and optimising resource use and data management processes.

This work has some aspects that add to the current corpus of research. First, to work with conceptual models in different languages as a previous step to integrate them. Second, from the point of view of geographical and cultural differences, this work contributes by aligning an authoritative data model of a National Spatial Data Infrastructure (NSDI), in this case, the Brazilian one, with a global collaborative data model, as is the OSM. Third, in the scope of the analysis and purpose of the application, this research contributes to adding collaborative data to the official mapping for updating and completeness. Finally, the method can be applied in other geographical contexts, even though the example of Brazil is used, and does not exclude the issue of generalisation versus specificity.

1.1 Background and Related Work

The emergence of Volunteered Geographic Information (VGI) has affected the geospatial community and society. VGI can provide real-time local knowledge data and has an ecosystem of open tools, data and dynamic communities that make it a valuable source of information. However, the heterogeneity of VGI is a significant challenge for government agencies considering its use for mapping updates (COLEMAN, 2013; FOODY et al., 2015; GIRRES and TOUYA, 2010).

From a semantic perspective, official datasets are reliable but often lack descriptive information outside of specific administrative and commercial purposes. Spatial Data Infrastructure (SDI) initiatives, such as the Infrastructure for Spatial Information in Europe (INSPIRE, 2023) and the National Spatial Data Infrastructure (INDE) in Brazil (INDE, 2023), address the data model heterogeneity and pursue interoperability. Schumacher (2021) in Germany and Tsiavos et al. (2012) in Greece proposed methods to harmonise national data with the European initiative with promising results. According to Mooney and Corcoran (2011), integrating VGI data is crucial for creating more accessible and user-centred SDIs.

The semantic integration of geospatial databases, such as OSM, with official databases using ontologies has seen some development. For example, Ludwig and Zipf (2019) studied the tags describing urban green areas in four cities and found that mapping context and purpose influence the depiction of vegetation on the OSM platform. Novack et al. (2018) proposed graph-based strategies to match OSM and Foursquare data points. Estima and Painho (2013) established a correspondence between CORINE Land Cover and OSM with remarkable results (76% accuracy). Al-Bakri and Fairbairn (2012) analysed the similarity between Ordnance Survey (OS) and OpenStreetMap data and found semantic mismatch due to automatic and generic semantic correspondence systems like WordNet. Finally, Cheatham et al. (2020) explored the alignment of hydrography domain ontologies and found that existing automated systems had limited success but they argued that manual alignment could be used to enhance information tasks.

In the context of the Semantic Web, Codescu et al. (2011) developed an ontology for OSM tags in English, while Ahmadian and Pahlavani (2022) evaluated the heterogeneity in the descriptions of the OSM tags and proposed a formal concept analysis to perform the integration of the building categories with the corresponding CityGML classification. Finally, Neumaier et al. (2018) integrated and linked datasets to create a base knowledge graph of geo-entities, adding semantic labels and disambiguating entities using context and hierarchy.

Despite efforts to align semantic data using OSM with SDI data, there is still a lack of standard practices. In addition, automatic tools for handling spatial data need improvement, especially when dealing with heterogeneous concepts across diverse languages and realities. Hence, manual alignment has been chosen in this work, and the contribution of meta-model matching is a step towards enhancing semantic alignment tools yet to come.

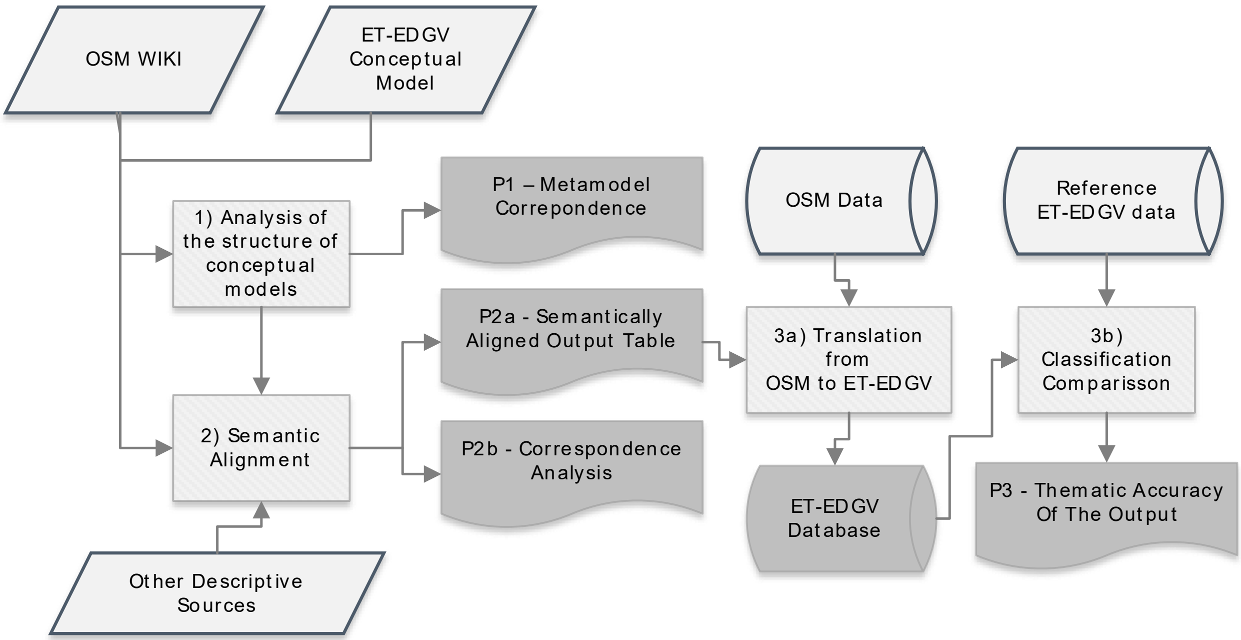
2 MATERIALS AND METHODS

Instruments are needed to achieve synergy between agents exchanging information represented by ontologies (and conceptual models), making them compatible and ensuring interoperability (CHEATHAM et al., 2020; FELICÍSSIMO and BREITMAN, 2004). According to Fridman and Musen (1999), there are two possible approaches: merging the ontologies to generate a single ontology or aligning the ontologies by establishing links between them to reuse each other's information. In the ontology's alignment case intended for this research, the ontologies are kept separate in an original way, increasing links between their equivalent concepts. This approach is vital because both conceptual models are existing standards, operated and updated by different entities. Therefore, it is possible to maintain interoperability over time by creating a shared and upgradeable alignment.

Thus, to perform the semantic alignment as background knowledge to integrate geospatial data of authoritative and collaborative mappings, we set out the method described in Figure 1, which consisted of three steps. In the first step, the structure of the two conceptual models is examined, resulting in a meta-model correspondence (P1). In the second step, the semantic alignment is performed based on the two conceptual models and other descriptive sources. This step produces two outputs: a semantically aligned output table (P2a) and semantic correspondences (P2b) analysis. The latter product includes identifying and classifying all possible relationships between elements of the input models and their multiplicities. This analysis also included a perspective on the relationship between the various types of geometries, a fundamental step when the focus is on geospatial data models. Finally, in the third step, the quality assessment is conducted, which consists of two processes. The first process is translating from OSM to ET-EDGV, resulting in an ET-EDGV database. The second process is a classification comparison between the database created in the last step and a reference ET-EDGV database, which then calculates the thematic accuracy of the output (P3). The thematic accuracy is an indicator of the overall quality of the alignment process.

We perform the semantic correspondence between the elements while documenting the process using natural and graphic language. First, we identified each data model's concepts, classes, relationships, axioms and instances and aligned the geometric and conceptual elements. Furthermore, we performed a quality assessment study case using semantic alignment for potential categories.

Figure 1– Methodological process and products.



Source: The authors (2024).

Two data models provided the core semantic information for alignment. The first one is the Technical Specification for Structuring Vector Geospatial Data (ET-EDGV 3.0) (CONCAR, 2023), the conceptual and semantic model of Brazilian terrestrial reference mapping documented in PDF format in Brazilian Portuguese. The National Cartography Commission, responsible for the National Cartographic System's standards, homologated the document. The Brazilian Spatial Data Infrastructure (INDE) also adopts this reference data standard. The ET-EDGV model is based on the Object Modelling Technique for Geographic Applications (OMT-G), which uses the primitives defined for the Unified Language (UML) class diagram, with the introduction of geographic primitives to expand the semantic representation (BORGES et al.,2001). In addition, the ET-EDGV contains class diagrams, tables with classes and attributes and their definitions, and domain lists. The OMT-G model was created in the late 1990s before ISO/TC211 established the ISO 19125:2004 standard (Geographic information — Simple feature access). Therefore, the definitions of spatial classes and relationships provided in the ET-EDGV standard are not exactly implementable in later spatial database management systems which follow the ISO standard. The Brazilian standard is, however, implemented in databases from topographic data producers such as IBGE, DSG and various municipalities, with some adjustments.

The second source was the OpenStreetMap conceptual model, based on Extensible Markup Language (XML) and published through a wiki (OPENSTREETMAP WIKI, 2023). We used the OSM wiki page in Brazilian Portuguese when available. The model has geometric data structures such as "nodes", "ways" and "relations". The semantic properties of each of the elements are "tags". They have two parts: a key and a value, defined in English. For example, a residential street should be semantically described by the tag "street = residential".

When there was doubt or inconsistencies about the alignment, we referred to other cartographic standards besides the two conceptual models (CONCAR, 2023; BRASIL,1998). For example, to obtain information about land routes and road infrastructure, we looked up national standards and technical manuals on the subject (BRASIL 1997, 2011; DNER, 1996; DNIT, 2006; DNIT, 2007, 2010, 2017, 2020). For trade, services, and industrial building information, we consulted specific manuals and legislation (IBGE, 2015; BRASIL, 2003). Additionally, we utilised dictionaries, wikis, images, and news sources in general.

3 RESULTS AND DISCUSSION

3.1 Analysis of the Structure of Conceptual Models

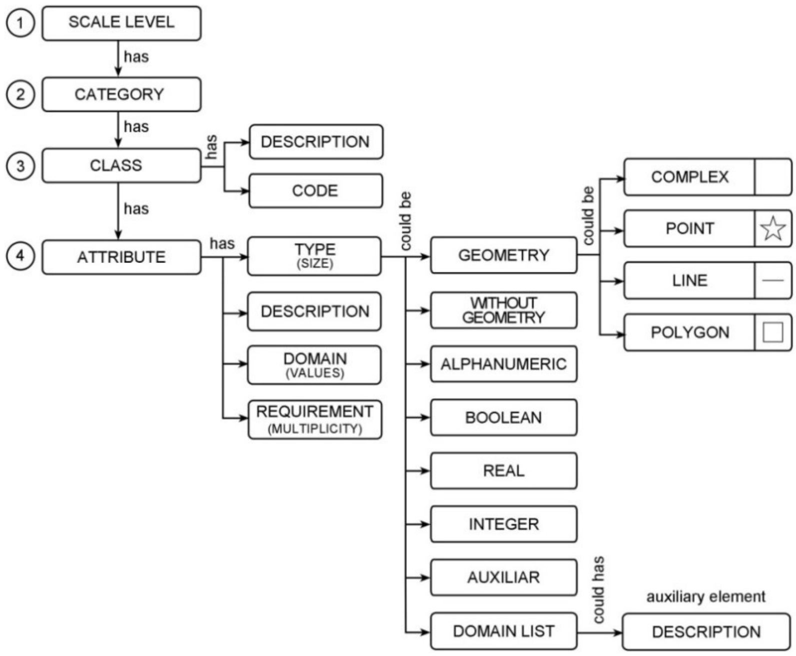
The first procedure analysed the conceptual structures of the two data models, ET-EDGV 3.0 and OpenStreetMap.One consequence of the fact that conceptual models have different structures is that the hierarchy of categories (in a generic way, an EDGV_Superclass), classes (EDGV_Class), attributes (EDGV_SubClass), and domain values (EDGV_SubSubClass) of ET-EDGV 3.0 do not directly match the keys (OSMClass) and tags (key+value) (OSMSubClass) of OpenStreetMap. The hierarchical structure of ET-EDGV 3.0's categories, classes, attributes, and domain values did not match the keys and tags in OpenStreetMap, leading to challenges in aligning the two models. These challenges included missing or unclear definitions, translation inconsistencies, and concepts with the same meaning but different names.

Three graphic representations were produced to understand better the two models' semantic and conceptual structures: a meta-model of ET-EDGV 3.0, a diagram of the OSM conceptual and semantic model structure, and a diagram showing possible correspondences between the structural and semantic elements of the two models. In addition, the graphical representations provided a better understanding of the hierarchical order and relationships between elements at each level.

The first graph, represented in Figure 2, depicts the meta-model of ET-EDGV 3.0, showcasing its conceptual and semantic structure. The Brazilian reference model is classified into small (MapTopoPE) and large (MapTopoGE) scales at the first taxonomic level. The second level comprises categories of information categorised by sets of classes or groups of features. The third taxonomic level includes the classes and provides a semantic description and code.

The fourth level includes attributes with a semantic description, domain values, and multiplicity. The attribute type includes the geometries representing the features, features without geometry, alphanumeric, boolean, real, integer, auxiliary, and domain list values. Additionally, semantic descriptions aid in understanding the meaning of domain lists. The geometry types could becomplex geometry, point, line or polygon. A class with complex-type geometry is one whose geometry may consist of more than one geometric primitive. This may occur in object classes where at least one instance has more than one geometric primitive or instances are represented by aggregating instances of classes of objects with different geometric primitives CONCAR (2018). This concept is similar tothe OGC geometry collection OPENGIS® (2011).

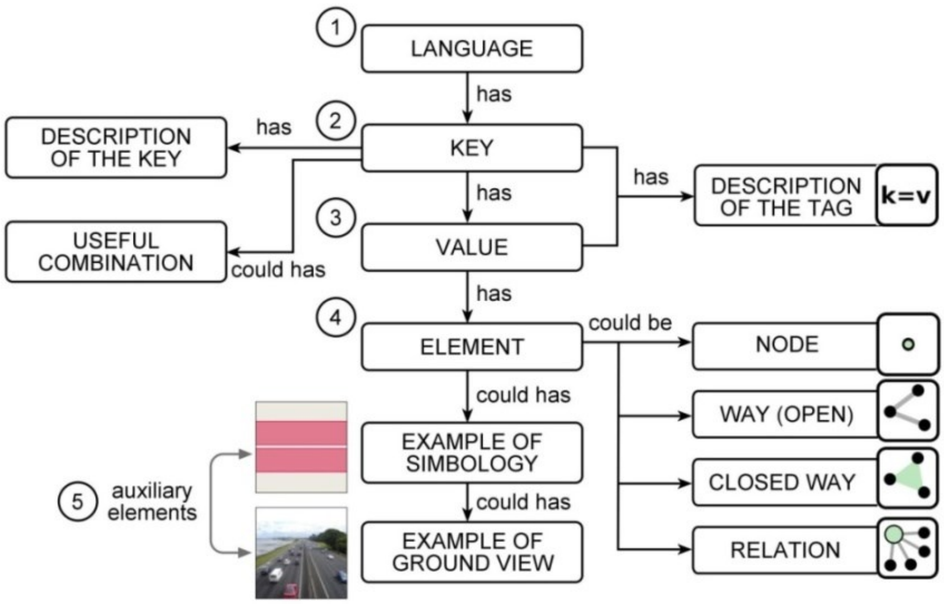
Figure 2 – The conceptual and semantic meta-model of ET-EDGV 3.0.



Source: The authors (2024).

Figure 3 shows the second graphical representation, the OpenStreetMap conceptual and semantic model structure diagram. At the first taxonomic level, language is classified since OSM is a global platform, and its wiki provides translations of semantic descriptions in various languages. However, keys and values remain in their original language, English. The second level categorises the key element, while the third level categorises the value element. The key has a general classification function, similar to the category in the Brazilian model. The combination of a key and a value has a classification function like the classes in ET-EDGV. In addition, its semantic description and "useful combination" are included for each key. The key description provides an overview of the tags composing it, while "useful combinations" provide hints of keys to be used with the main key + value pairs. The fourth level refers to the term element, which indicates the possible geometry types for the feature described by the tag, including node, way, closed way and relation. The conceptual structure of OSM also includes auxiliary elements, such as the example of symbology and a ground view, to aid in understanding the model's semantic descriptions.

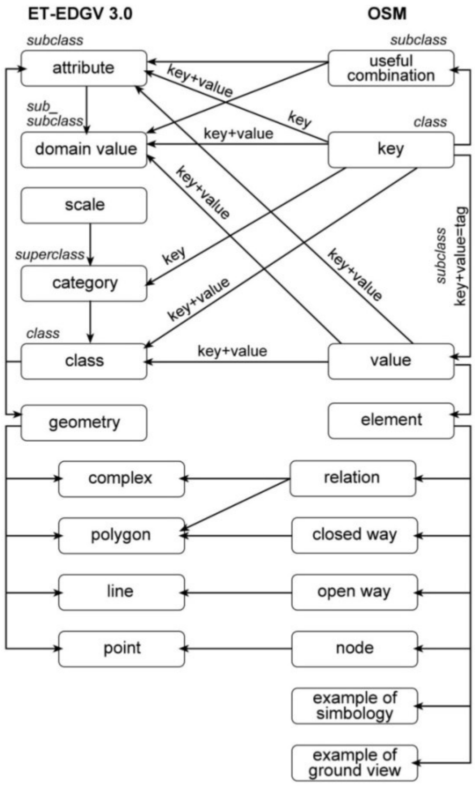
Figure 3 – Diagram of the OpenStreetMap conceptual and semantic model.



Source: The authors (2024).

Figure 4 shows the third graphic representation resulting from the study of conceptual models' structure and presents the possible correspondences between the structural and semantic elements of ET-EDGV 3.0 and OSM conceptual models. For semantic elements, it is possible to observe that an OSM key (OSM_Class) can correspond to a category (in a generic way, EDGV_SuperClass) or attribute (EDGV_SubClass) in ET-EDGV 3.0. Likewise, a key+value (OSM_SubClass) can match an attribute (EDGV_SubClass), a domain value (EDGV_SubSubClass), or a class (EDGV_Class). Likewise, useful combinations (OSM_SubClass) can match an attribute (EDGV_SubClass) or domain value (EDGV_SubSubClass). Concerning geometry, the area element (closed way) corresponds to the polygon element in ET-EDGV 3.0. The way element (open way) corresponds to the line element. The node element corresponds to the point element. The relation elementmatches the complex geometry and the area element in ET-EDGV 3.0.

Figure 4 – Correspondence between the conceptual elements of ET-EDGV 3.0 and OpenStreetMap.



Source: The authors (2024).

Studying conceptual models and their synthesis through the representations above allowed us to understand their semantic and conceptual structures and their differences and similarities in the semantic alignment process systematisation.

3.2 Semantic Alignment

The semantic alignment was made from the OSM to ET-EDGV 3.0, first treating the keys and then each of the key+value pairs. Only the standard key values presented on the OSMwiki were processed in the alignment. The 5-step method for performing semantic alignment from OSM to ET-EDGV 3.0 includes the five next criteria:

a) manual assessment of semantic similarity to compare the concepts and features to determine whether they are equal, synonyms or similar;

b) examine the words, synonyms, or similar descriptions referring to the terms in the conceptual models if the first selection requirements are not met;

c) textual analysis of the definitions of the concepts in both sources if the first and second selection

 DOI: http://dx.doi.org/10.14393/rbcv76n0a-72070

criteria are not met or partially met;
d) consultation of other sources, including images, until similar or equivalent meaning;
e) search by the corresponding category, class, attribute, or domain value in ET-EDGV 3.0.

Additionally, if an ET-EDGV element does not match the analysed key but could have a corresponding key in OSM, other keys are searched to find a match.

Some classes were aligned directly through the semantic descriptions of the model, such as the class Arvore_Isolada (isolated tree) of ET-EDGV 3.0 and the key "natural = tree". However, in some cases, such as bicycle parking (OSM tag "amenity = bicycle_parking"), the alignment required more in-depth analysis, as this feature does not yet exist in ET-EDGV 3.0. It was necessary to look for other sources, such as the Curitiba city hall website, which has definitions of the elements of thecycling structure of the city, the OSM wiki, which providesthesemantic descriptions of tags and images on the internet toestablish the meaning of the feature. In this case, we aligned the tag amenity = bicycle_parking (OSMSubClass) with the domain value “Others” (EDGV_SubSubClass) of the attribute “Purpose of Use” (EDGV_SubClass) in the class “Parking” (EDGV_Class) of the “Transport System” category (EDGV_Superclass) of the Brazilian data model. In addition, during the alignment, terminological issues emerged regarding certain terms, such as "estrada" (motorway) and "rodovia" (road), which are used synonymously in Brazil. Although the ET-EDGV 3.0 descriptions are different for "road" and "highway" they are not enlightening regarding the difference between the two terms. After analyzing the descriptions of the corresponding tags in OSM, the situation was the same. Therefore, technical documents from the National Department ofLand Infrastructure (DNIT), National Department of Roads and Highways (DNER) and the Brazilian Traffic Code (CTB) were consulted. The Brazilian Traffic Code (CTB) differentiates “road” as “unpaved rural road” and “highway” as “paved rural road”. The DNIT Glossary of Road Technical Terms (IPR-700) provides definitions for these terms that contradict each other. Thus, "road" and "highway" are two technical terms used as synonyms by domain experts, demonstrating that further work, especially in semantics and ontologies, is needed to clarify certain concepts. Among these definitions, the fact that the IPR also included words in other languages was of great value in the semantic alignment of these two terms. However, this type of description is ambiguous, a situation that generates problems in semantic alignment, as was found in the tests carried out.

The application of semantic alignment produced three results. This documentation has been made available in the GitHub repository (Github, 2023) and in the Federal University of Paraná (UFPR) Scientific Database (BDC) (UFPR, 2023), with an associated Digital Object Identifier (DOI) to facilitate consultation, sharing and citation.

3.2.1 SEMANTIC ALIGNMENT OUTPUT TABLE

We produced semantically aligned output tables as the primary result, which comprises 4810 concept alignments organised into 18 tables for each category in ET-EDGV (including Green Areas, Urban Mobility, Buildings, Transport, Vegetation, Sanitation, Topography, Hydrography, Localities, Borders, Economic Structure, Energy, and Communications). The tables are available through the link (http://dx.doi.org/10.5380/bdc/53) and present alignment in Portuguese by dividing the ET-EDGV subdivisions (MapTopGE and MapTopPE) and for each category of ET-EDGV with the primary associated OSM key. We also calculated and analysed the percentage of classes of each ET-EDGV 3.0 aligned with the corresponding OpenStreetMap keys, enabling the identification of categories with the most significant potential for integration between the databases. These tables are structured according to the ET-EDGV 3.0 hierarchy, providing ET-EDGV classes, attributes and domain values, and the OSM model's corresponding tags (keys+values) with respective semantic descriptions. Chart 1 is an example of one of these tables, translated to English.

Due to differences in the structure of each conceptual model, not all elements match. For example, in some cases, only the class or attribute is aligned, or one or more domain values are aligned. Hence, if the table field is not filled, no corresponding OSM key or value will be found for that element in ET-EDGV 3.0.

 DOI: http://dx.doi.org/10.14393/rbcv76n0a-72070

Chart 1 – ET-EDGV 3.0 - Transport/Roadway System Category OSM – Key Highway
The key highway=* is used to identify any type of road, street, or path. The value of the key helps indicate the importance of the highway within the road network as a Whole (to be continued).

ET-EDGV 3.0						OpenSTreetMap (OSM)		
EDGV class	**Class description**	**Attribute**	**Attribute description**	**Value Domain**	**Value Domain Description**	**Key**	**Value**	**Tag (Key+Value) Description**
Via G= complex geommetry	**Via** is the land transit route, intended for motor vehicles, except the carriage way and those belon-ging to the Railway System.					**highway**	*	The key **highway=*** is the main key used for identifying any kind of road, street or path. The value of the key helps indicate the importance of the highway within the road network as a whole.
Via G= complex geommetry	**Via** is the land transit route, intended for motor vehicles, except the carriage way and those belonging to the Railway System.	**typeOfVia**	Indicates the type of travel route.	**1) Motorway**	Fast traffic lane, with all access controlled, without level crossing and intended exclusively for motor vehicles, with solid surface (asphalt, concrete or paving), with a minimum of four lanes, with physical separa-tion between traffic lanes, representable scaled or not.	**highway**	**motorway**	**highway=motorway**: A restricted access major divided highway, normally with 2 or more running lanes plus emergency hard shoulder. Equivalent to the Freeway, Autobahn, etc..
				2)Public place	Any public space for circulation of people, vehicles and goods, such as avenues, streets, squares, gardens, parks, etc., recognized by the administration of a municipality or by the local community, associated with a name that is generally known.	**highway**	**unclassified**	**highway=unclassified**: The least important through roads in a country's system – i.e. minor roads of a lower classification than tertiary, but which serve a purpose other than access to properties. (Often link villages and hamlets.) The word 'unclassified' is a historical artefact of the UK road system and does not mean that the classification is unknown; you can use highway=road for that.
						highway	**residential**	**highway=residential**: Roads which serve as an access to housing, without function of connec-ting settlements. Often lined with housing.
						highway	**living_street**	**highway=living_street**: For living streets, which are residential streets where pedestrians have legal priority over cars, speeds are kept very low and where children are allowed to play on the street.

Source: The authors (2024).

 DOI: http://dx.doi.org/10.14393/rbcv76n0a-72070

Chart 1 – ET-EDGV 3.0 - Transport/Roadway System Category OSM – Key Highway

The key highway=* is used to identify any type of road, street, or path. The value of the key helps indicate the importance of the highway within the road network as a whole (conclusion).

ET-EDGV 3.0						OpenStreetMap (OSM)		
EDGV class	**Class description**	**Attribute**	**Attribute description**	**Value Domain**	**Value Domain Description**	**Key**	**Value**	**Tag (Key+Value) Description**
Via G= complex geommetry	**Via** is the land transit route, intended for motor vehicles, except the carriage way and those belonging to the Railway System.	**typeOfVia**	Indicates the type of travel route.			**highway**	**service**	**highway=service**: For access roads to, or within an industrial estate, camp site, business park, car park, alleys, etc. Can be used in conjunction with service=* to indicate the type of usage and with access=* to indicate who can use it and in what circumstances.
						highway	**pedestrian**	**highway=pedestrian**: For roads used mainly/exclusively for pedestrians in shopping and some residential areas which may allow access by motorised vehicles only for very limited periods of the day.
				3)Highway	Road intended mainly for the traffic of vehicles equipped with tires, which crosses a territorial extension, connecting two or more locations.	**highway**	**trunk**	**highway=trunk**: The most important roads in a country's system that aren't motorways. (Need not necessarily be a divided highway.)
						highway	**primary**	**highway=primary**: The next most important roads in a country's system. (Often link larger towns.)
						highway	**secondary**	**highway=secondary**: The next most important roads in a country's system. (Often link towns.)
						highway	**tertiary**	**highway=tertiary**: The next most important roads in a country's system. (Often link smaller towns and villages).
(...)								
		typeOf Paving	Indicates the type of structure built after earthworks through layers of various materials with different resistance and deformability characteristics.	**1) Asphalt**	Pavement built with asphalt, that is material of variable consistency, dark brown or black, and in which the predominant constituent is bitumen, which may occur in nature in deposits or be obtained by refining Petroleum.	**surface**	**asphalt**	**surface=asphalt:** Short for asphalt concrete - mineral aggregate bound by asphalt. Most such features are tagged as paved without specifying exact surface.
(...)								

Source: The authors (2024).

Out of the 19 categories of ET-EDGV 3.0, 69 OSM keys were aligned, except for the Reference Points category. Table 1 shows the total ET-EDGV 3.0 classes aligned in each category. Categories such as Limits and Localities, Vegetation, Relief, Hydrography, and Transport/Waterway System had low alignment percentages, as they require specific technical knowledge or are dependent on legislation. However, the categories with a high percentage of alignment, such as Urban Mobility Structures, Buildings, and Green Areas, are composed of artificial features, mainly in urban areas. These results suggestthat categorisation requires less abstract reasoning when the characteristics of the real world are easily recognised (TVERSKY and HEMENWAY, 1984). Furthermore, taxonomic classification mechanisms were observed in bothsources' knowledge structures, initially analysing the parts and attributes of each category and then grouping them according to similarity. It was also observed that knowledge structures are made through the taxonomic classification mechanism in the conceptual models of both sources. This information organisation demonstrates that the sets of categories were initially analysed through their parts and other attributes and subsequently grouped by types according to their similarity (TVERSKY and HEMENWAY, 1984).

Table 1 – Total ET-EDGV 3.0 classes aligned in each category.

Topographic Mapping on Large Scales		
Category ET-EDGV 3.0	**Total Classes Aligned**	**# OSM Keys**
Green Area	3 of 4 (75%)	2
Large Scale Topographic Mapping Base Classes	13 of 29 (45%)	22
Culture and Leisure	5 of 9 (56%)	10
Buildings	28 of 35 (80%)	28
Urban Mobility Structure	8 of 9 (89%)	11
Topographic Mapping on Small Scales		
Energy and Communications	11 of 16 (69%)	4
Economic Structure	2 of 3 (67%)	3
Hydrography	8 of 20 (40%)	3
Boundaries and Localities	4 of 25 (16%)	3
Relief	7 of 18 (39%)	1
Basic Sanitation	2 of 4 (50%)	3
Transportation System	10 of 16 (63%)	25
Transport/Airport System	2 of 2 (100%)	1
Transport System/ Pipelines	2 of 2 (100%)	2
Transport/Rail System	5 of 5 (100%)	15
Transport System/ Waterway	2 of 7 (29%)	2
Transport/Road System	2 of 3 (67%)	5
Vegetation	7 of 13 (54%)	2

Source: The authors (2024).

3.2.2 CORRESPONDENCE ANALYSIS

The second result of the alignment between the ontologies (http://dx.doi.org/10.5380/bdc/52) is a table that presents the possible semantic correspondence cases with their respective multiplicities and the potential instances of geometric correspondence between the conceptual models. Table 2 shows, as an example, an excerpt from this table translated into English. The results are divided by the ET-EDGV subdivisions (MapTopGE and MapTopPE), represented in the first line of the table. The categories of information, the EDGV_SuperClass, for example, Transport/Road System, are shown in the second line. The semantic and geometric elements of the two conceptual models with possible cases of correspondence are the headers in the third and fourth lines. The feature geometries described in the ET-EDGV column "G" can be A=area; P=point; L=line; C=complex geometry; G=generic; and conv=conventional. In the case of OSM, the geometries in column "G" can be: a=area (closed path); c=line (way); n=point (node) and; r=relation.

 DOI: http://dx.doi.org/10.14393/rbcv76n0a-72070

Table 2 – Excerpt of the table of the semantic correspondence classification.

1. Topographic Mapping In Small Scales									
1.13 Transport / Road System									
ET-EDGV 3.0				**OSM**			**Correspondence**		
Class	**Attribute**	**Domain value**	**G**	**Key**	**Value**	**G**	**Semantic**	**Multiplicity**	**Geometric**
Via	-	-	G	highway	*	n,c	1	m:1	6
	typeOfVia	1) Motorway	G	highway	motorway	c	3	m:1	6
		2) Public place	G	highway	unclassified	c	3	m:1	6
		2) Public place	G	highway	residential	c	3	m:1	6
		2) Public place	G	highway	living_street	c	3	m:1	6
		2) Public place	G	highway	service	a, c	3	m:1	6
		2) Public place	G	highway	pedestrian	a, c	3	m:1	6
		3) Highway	G	highway	trunk	c	3	m:1	6
		3) Highway	G	highway	primary	c	3	m:1	6
		3) Highway	G	highway	secondary	c	3	m:1	6
		3) Highway	G	highway	tertiary	c	3	m:1	6
		4) Link between lanes	G	highway	motorway_link	c	3	m:1	6
		4) Link between lanes	G	highway	trunk_link	c	3	m:1	6
		4) Link between lanes	G	highway	primary_link	c	3	m:1	6
		4) Link between lanes	G	highway	secondary_link	c	3	m:1	6
		4) Link between lanes	G	highway	tertiary_link	c	3	m:1	6
		5) Junction stretch	G	highway	motorway_link	c	3	m:1	6
		5) Junction stretch	G	highway	trunk_link	c	3	m:1	6
		5) Junction stretch	G	highway	primary_link	c	3	m:1	6
		5) Junction stretch	G	highway	secondary_link	c	3	m:1	6
		5) Junction stretch	G	highway	tertiary_link	c	3	m:1	6
		6) Others	G	highway	track	c	3	1:1	6
		6) Others	G	highway	bus_guideway	c	3	1:1	6
		6) Others	G	highway	road	c	3	1:1	6
		6) Others	G	highway	footway	c	3	m:1	6
	coating	-	G	surface	*	a, c	2	m:1	6
	coating	1) Uncoated	G	surface	unpaved	a, c	3	1:m	6
	coating	1) Uncoated	G	surface	dirt	a, c	3	1:m	6
	coating	1) Uncoated	G	surface	ground	a, c	3	1:m	6
(...)									

Source:The authors (2024).

The last three columns of the table show the classification of possible matching cases, either in geometric or semantic terms, as well as their multiplicity. Additionally, the classes are described in Chart 2. Thus, this analysis allows an overview of the alignment concerning the aspects mentioned. Furthermore, this chart shows the general rules obtained in this research to apply semantic alignment to any other sources, schemas, or models.

Chart 2 – Classification of possible matching cases.

Semantic Matching Cases	1= class/key+value
	2= attribute/key+value
	3= domain/key+value
	4= attribute/domain which is in the OSM and is not in the ET-EDGV
Multiplicityof Semantic Matching Cases	1:1= exact matches
	m:1= one key+value is equivalent to more than one domain, attribute and/or class
	1:m= more than one key+value is equivalent to a domain, attribute or class
Geometric Types Matching Cases	5= same geometry
	6= complex geometry (ET-EDGV) /relation (OSM)
	7= different geometry: more complex/detailed geometry in OSM
	8= different geometries: more complex/detailed geometry at ET-EDGV
	0= conventional class/no geometry

Source:The authors (2024).

When dealing with geospatial databases, comparing geometry types is a crucial aspect to consider. First, it is essential to note that OpenStreetMap (OSM) contains data with varying levels of detail, depending on the available inputs and the contributor's interpretation. For example, "amenity=school" may be associated with an area tagged "landuse=education" or with a node. On the other hand, the scales of the reference mapping are consistent in their level of detail, meaning that all schools are points at one scale and polygons at another. Thus, we may not always compare similar geometry types when integrating data. For example, the school building mentioned in OSM may be integrated into the official base as a less complex geometry, perhaps its centroid (point). These additional conversion and harmonisation procedures must be considered when automating the integration process.

3.2.3 THEMATIC ACCURACY CASE STUDIES

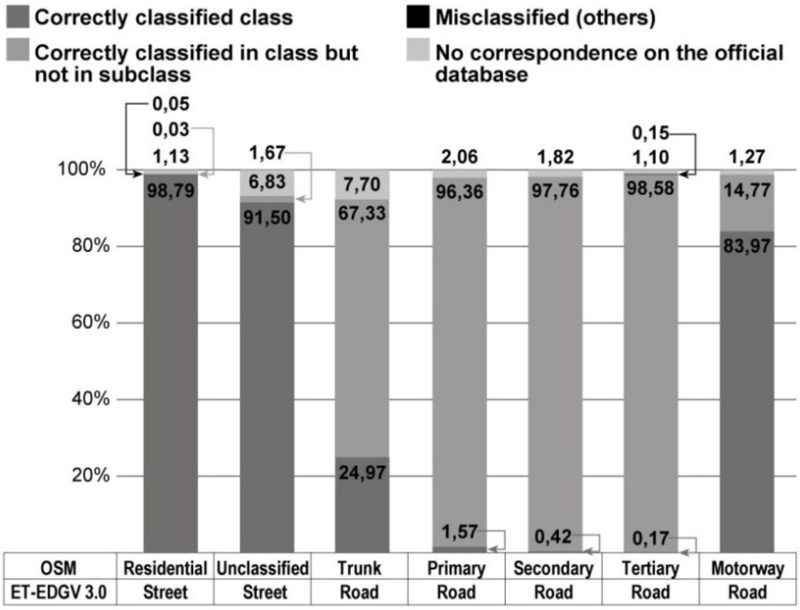
In order to evaluate the semantic alignment results and demonstrate how the alignment performs when applied to real-world data, we conducted case studies between two datasets in different areas. The first case study area, in the urban area of Curitiba, state of Paraná (PR), featured large-scale reference data from the Institute for Research and Urban Planning of Curitiba (IPPUC), with production dates ranging from 1972 to 2000 and scales of 1:5.000, 1:2.000 and 1:1.000. The second case study area was carried out for a rural and urban area in the municipality of Campos dos Goytacazes, State of Rio de Janeiro (RJ), using the continuous dataset of the State of Rio de Janeiro produced in 2018 by the Brazilian Institute of Geography and Statistics (IBGE) at a scale of 1:25.000.

The two reference databases had to be made compatible with ET-EDGV 3.0 since Curitiba's database uses its own conceptual model, and Campos dos Goytacazes' database is available with ET-EDGV 2.1.3 as a model. However, this conversion was done without additional semantic discrepancies. The "highway" key data, which corresponded to the via_deslocamento and trecho_rodoviário classes in the Brazilian conceptual model by the semantic alignment, were imported. The homologous features in the two datasets were first identified by comparing the OSM vectors with the positional uncertainty of the corresponding lines in the official mapping. The features from OSM were then compared with those in the official database to identify those in both databases. This procedure separated our results into features that matched the official data and those that did not. The matching records were then used to calculate the thematic accuracy.

The 15-metre buffer was applied according to an estimated accuracy for OSM data between 1:20.000 and 1:30.000 in Brazil. This procedure only included OSM features within this threshold compared to the reference base data. For each OpenStreetMap segment, a spatial attribute union was applied to identify the reference classification of that segment. The thematic accuracy of data imported from OSM was measured by registering the number of correctly classified features according to the producers of the official mapping. Finally, the calculated accuracy is the percentage of data imported via alignment with the correct classification compared to the official one. The study case results were compiled into two graphs showing the percentages of OSM features sorted correctly and incorrectly.

For Curitiba, the best matching percentages were obtained for the "highway=residential" tag (98.79%) and the "highway=motorway" tag (83.97%). For the tag "highway=unclassified", the accuracy was 91.50%. However, OSM's semantic description for this tag is ambiguous, resulting in an unresolved match. Concerning the misclassifications, which only agreed on the class but not on the subclass, the results were as follows: "highway=trunk" tags (67.33%), "highway=primary" (96.36%), "highway=secondary" (97.76%) and "highway=tertiary" (98.58%). There are several potential factors for the lack of 100% thematic accuracy in these cases. Several factors may have led to the low percentage of correctly classified segments for specific tags, such as "highway=unclassified", "highway=secondary", and "highway=tertiary". One of these factors is the ambiguous semantic descriptions of these tags, which have been aligned multiple times with different classes and domain values (multiplicity 1:m in Table 1). Another factor is that, in addition to alignment, a context analysis may be required for a system to match automatically. For example, in the case of Curitiba, several highways that have passed to municipal management are still perceived as roads, which may have influenced the results. Additionally, misclassifications by the OSM contributors are also a possibility. The procedures described previously calculated the thematic accuracy of the alignments for the category "highway". For Curitiba, large-scale data from the city hall obtained an overall accuracy of 69.2% of the 49.808 corresponding segments identified in the two bases. However, this value varied significantly for each class. Figure 5 shows Curitiba's semantic alignment evaluation test results.

Figure 5 – Alignment quality assessment study case for Curitiba (PR).

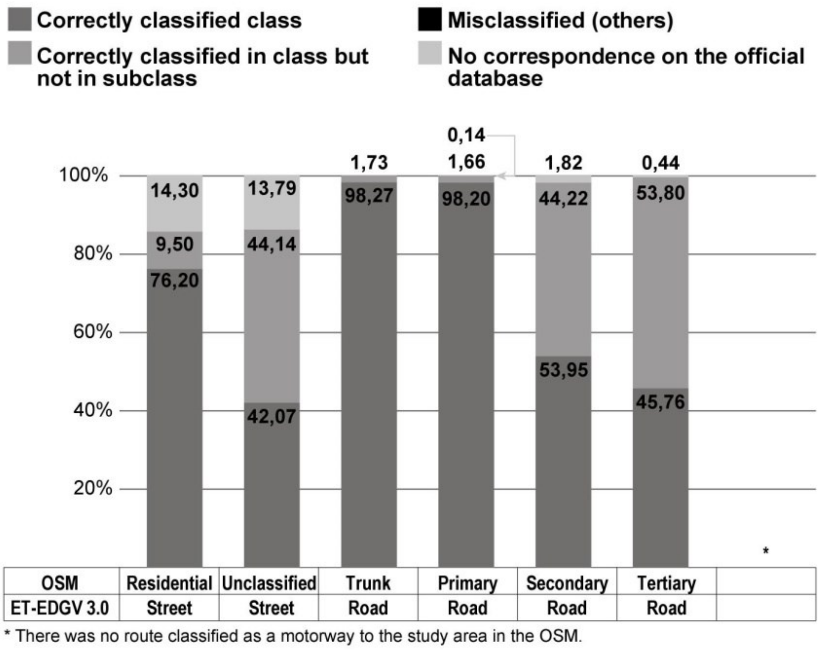


Source: The authors (2024).

Figure 5 shows a small percentage of incorrect classification after the alignment. One hypothesis is that there are original misclassifications by OSM contributors. It is worth noting that there are OSM features without correspondence in the official database, such as internal streets of condominiums that are not part of the official mapping or non-existent streets in the official database, either due to a lack of updating or omissions in the mapping process.

Figure 6 shows the results of the second study case performed for Campos dos Goytacazes in Rio de Janeiro State, including urban and rural areas.

Figure 6 – Alignment quality assessment study case for Campos dos Goytacazes (RJ).



Source: The authors (2024).

The overall thematic accuracy calculated was 68.6%. The best percentage of segments correctly classified by alignment was for "highway=trunk" tags (98.27%) and "highway=primary" (98.20%), corresponding to road and highway segments, and for the "highway=residential" tag (76.20%), corresponding to sections of residential streets. The improvement in the percentages of the first two tags' classification compared to the Curitiba test rates is due to the area of Campos dos Goytacazes being more extensive and including both urban and rural areas. The low percentage of correctly classified segments for the other tags can be attributed to factors such as ambiguous semantic descriptions, the need for local context analysis, and possible errors in the original data classification by the OSM contributors.

Regarding the study cases, a variation in categorisation occurs regarding the OSM tags' values aligned with the corresponding domain values in ET-EDGV 3.0. In the case of values with the best percentage in alignment, for example, residential, unclassified and motorway, it is analysed that these terms are at the level of basic abstraction, demanding a type of intermediate reasoning between the abstract and the concrete. The features (a street and a road) are easily differentiated at this level, reflecting the real world's structure. On the other hand, in the case of the values that had the worst percentage in the alignment, for example, trunk, primary, secondary and tertiary, it is observed that, although they are at the same hierarchical level of the taxonomy of ET-EDGV 3.0, they involve a second hierarchisation in the OSM, that of the highways themselves. In this case, the type of reasoning demanded in the mental categorisation of these features is more concrete and characteristic of the lower level of abstraction, whose categorisation level within the taxonomy is more specific. In this respect, it is important to consider what additional attributes would be added to the highways to differentiate between primary, secondary and tertiary highways in mental categorisation. This categorisation also has low suggestion validity because it generally shares many attributes with highways or roads (TVERSKY and HEMENWAY, 1984).

In light of these concepts and analyses, it is possible to affirm that semantic alignment was more effective for features whose abstraction level is basic, requiring a degree of intermediate reasoning between concrete and abstract. Those categories allow the differentiation of features closer to what is seen in the real world: a road segment different from a building or a green area. The semantic alignment in the categorisations that went beyond this level of abstraction, whether more abstract or more concrete (notwithstanding other factors such as the ambiguity of semantic descriptions), requires additional analysis of the local context and aligned features.

3.2.4 GUIDELINES APPLYING THE METHOD IN OTHER CONTEXTS

The experience in creating and documenting the alignment and the quality analysis using real data enables us to list some recommendations for subsequent similar applications. The recommendations for integrating OpenStreetMap data into national mapping agencies or other geospatial data producers include:

a) identifying categories of the conceptual model with the most significant potential for integration, considering the specificities of the region or context;

b) making the conceptual models compatible by identifying relationships between categories, classes, attributes, and data geometry;

c) ensuring semantic compatibility between the elements and data of the conceptual models through comprehensive search and alignment, formalisation of ontologies, and generation of new tools;

d) it is also recommended that the material be maintained and made available collaboratively with mapping stakeholders and that correspondence quality tests be performed to ensure accuracy and promote adjustments;

e) in cases where the classifications are more detailed, we suggest analyzing the context of data use for thedata integration;

f) in cases where there are possible errors in the original data classification by the OSM contributorsis recommended a local context analysis.

4 CONCLUSION

This article discusses the alignment of the conceptual models of OpenStreetMap (OSM) and the Technical Specification for Structuring Vector Geospatial Data (ET-EDGV 3.0) to enable semantic and logical interoperability between authoritative and collaborative mappings. The method's results offer guidelines for integrating OSM data into other databases or vice versa, especially for national mapping agencies in other countries. The research covers all taxonomic levels of conceptual models and highlights categories of data with greater and lesser potential for integration. The alignment generated effectively matched classes at the upper and basic abstraction levels, while the lower or subordinate taxonomic levels required additional analysis of the local context and aligned features. The article concludes by suggesting future research on scale issues in conceptual models, the semantic generalisation of concepts, linguistic regionalisms in multiple languages, and legal and ethical issues regarding integrating official and collaborative data.

Acknowledgments

To the PPGCG UFPR (Postgraduate Program in Geodetic Sciences at the Federal University of Paraná) and the CNPq (National Council for Scientific and Technological Development) Brazil, grant number 141073/2017-9 and CAPES (Coordination for the Improvement of Higher Education Personnel), Brazil for the doctoral scholarship received by the main author.

Authors' Contribution

Conceptualisation, A. M. and S. C.; Methodology, A. M. and S. C.; Software, S. C.; Validation, S. C.; Formal Analysis, A. M. and S. C.; Investigation, A. M.; Writing-Original Draft Preparation, A. M.; Writing-Review & Editing, A. M.; Supervision, S. C.; All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

AHMADIAN, S.; PAHLAVANI, P. Semantic integration of OpenStreetMap and CityGML with formal concept analysis. **Transactions in GIS**, v. 26, n. 8, p. 3349–3373, 2022.

AL-BAKRI, M.; FAIRBAIRN, D. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. **International Journal of Geographical Information Science**, v. 26, n. 8, p. 1437–1456, 2012. DOI: https://doi.org/10.1080/13658816.2011.636012.

BORGES, K. A. V.; DAVIS, C. A.; LAENDER, A. H. F. OMT-G: An Object-Oriented Data Model for Geographic Applications. **GeoInformatica**, v. 5, n. 3, p. 221–260, 2001. DOI: https://doi.org/10.1023/A:1011482030093.

BRASIL. Lei 9.503 de 23 de setembro de 1997. **Institui o Código de Trânsito Brasileiro**. 7ª Ed., Brasília, DF. 1997.

BRASIL. Lei Complementar 116 de 31 de julho de 2003. Dispõe sobre o Imposto Sobre Serviços de Qualquer Natureza, de competência dos Municípios e do Distrito Federal, e dá outras providências. **Diário Oficial da União,** Brasília, DF. Seção 1, p. 3. 2003.

BRASIL. Lei 12.379 de 6 de janeiro de 2011. Dispõe sobre o Sistema Nacional de Viação – SNV; e dá outras providências. **Diário Oficial da União,** Brasília, DF. 2011.

BRASIL. EXÉRCITO BRASILEIRO. **Manual Técnico T 34-700 Convenções Cartográficas (1ª parte) Normas para o Emprego dos Símbolos**, 2ª edição, 1998.

CAMBOIM, S. P.; BRAVO, J. V. M.; SLUTER, C. R. An investigation into the completeness of, and updates to, the Open Street Map data in a heterogeneous area in Brazil. **ISPRS International Journal of Geo-Information**, v. 3, n. 4, p. 1366–1388, 2015.

CHEATHAM, M.; VARANKA, D.; ARAUZ, F.; ZHOU, L. Alignment of surface water ontologies: a comparison of manual and automated approaches. **Journal of Geographical Systems**, v. 22, n. 2, p. 267–289, 2020. DOI: https://doi.org/10.1007/s10109-019-00312-3.

CODESCU, M.; HORSINKA, G.; KUTZ, O.; MOSSAKOWSKI, T.; RAU, R. OSMonto - An Ontology of OpenStreetMap Tags. State of the Map Europe (SOTM-EU) 2011. **Anais.** 2011.

COLEMAN, D. J. Potential Contributions and Challenges of VGI for Conventional Topographic Base-Mapping Programs. In: D. Sui; S. Elwood; M. Goodchild (Orgs.), **Crowdsourcing geographic knowledge: Volunteered geographic information (VGI) in theory and practice**, Dordrecht, Netherlands, 2013. p.245–263.

CONCAR. Comissão Nacional de Cartografia. **Especificação Técnica para a Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV)** v. 3.0, 2018.

CONCAR.Comissão Nacional de Cartografia. **Especificação Técnica para a Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV)** v.3.0, 2018.

DNER. Departamento Nacional de Estradas de Rodagem. **Manual de Pavimentação (IPR-697)**. 2ª ed. ed. Rio de Janeiro, Brasil, 1996.

DNIT.Departamento Nacional de Infraestrutura Terrestre. **Manual de Pavimentação (IPR-719)**. 3ª ed. ed. Rio de Janeiro, Brasil, 2006.

DNIT. Departamento Nacional de Infraestrutura Terrestre. **Terminologias Rodoviárias Usualmente Utilizadas**. v. 1.1,Rio de Janeiro, Brasil, 2007.

DNIT.Departamento Nacional de Infraestrutura Terrestre. **Manual de Implantação Básica de Rodovia (IPR-742)**. 3ª ed. Rio de Janeiro, Brasil, 2010.

DNIT.Departamento Nacional de Infraestrutura Terrestre. **Glossário de Termos Técnicos Rodoviários (IPR-700)**. 2ªed. ed. Rio de Janeiro, Brasil, 2017.

DNIT.Departamento Nacional de Infraestrutura Terrestre. **Nomenclatura das Rodovias Federais**. Rio de Janeiro, Brasil, 2020.

DORN, H.; TÖRNROS, T.; ZIPF, A. Quality Evaluation of VGI Using Authoritative Data - A Comparison with Land Use Data in Southern Germany. **ISPRS International Journal of Geo-Information**, v. 4, n. 3, p. 1657–1671, 2015.

ELIAS, E. N. N.; FERNANDES, V. DE O. Qualidade dos Dados Geoespaciais do OpenStreetMap para os indicadores de Acurácia Posicional, Acurácia Temática e Completude. **Geografia (Londrina)**, v. 30, n. 2, p. 255–275, 2021.

ESTIMA, J.; PAINHO, M. Exploratory analysis of OpenStreetMap for land use classification. Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information. **Anais.** GEOCROWD '13, New York, NY, USA, 2013. p. 39-46. DOI: https://doi.org/10.1145/2534732.2534734.

FELICÍSSIMO, C. H.; BREITMAN, K. K. Uma estratégia para o alinhamento taxonômico de ontologias. I Workshop de Web Semântica, XVIII Simpósio Brasileiro de Engenharia de Software. **Anais.** v. 123, Brasília, DF, Brasil, 2004. Available in: http://www.ucb.br/ucbtic/WWS2004/. Accessed on: Apr 05, 2024.

FOODY, G. M.; SEE, L.; FRITZ, S.; et al. Accurate Attribute Mapping from Volunteered Geographic Information: Issues of Volunteer Quantity and Quality. **Cartographic Journal**, v. 52, n. 4, p. 336-344, 2015. DOI: https://doi.org/10.1080/00087041.2015.1108658.

FRIDMAN, N.; MUSEN, M. SMART: Automated Support for Ontology Merging and Alignment.In **Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management** (KAW), p. 1–2,1999. Available in: https://dl.acm.org/doi/proceedings/10.5555/645361. Accessed on: Apr 05, 2024.

GIRRES, J.-F.; TOUYA, G. Quality Assessment of the French OpenStreetMap Dataset. **Transactions in GIS**, v. 14, n. 4, p. 435–459, 2010.

GITHUB. Available in: https://github.com/Labgeolivre-UFPR/IntegracaoOSM_EDGV. Accessed on: Feb 02, 2023.

GRUBER, T.; WILLBERG, M. Signal and error assessment of GOCE-based high resolution gravity field models. **Journal of Geodetic Science**, v. 9, n. 1, p. 71–86, 2019.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Classificação Nacional de Atividades Econômicas (CNAE) Subclasses, v. 2.2.** Rio de Janeiro, RJ, 2015.

INSPIRE. European Comission. Available in: https://inspire.ec.europa.eu/. Accessed on: Feb 13, 2024.

LIMA, P.; CÂMARA, G.; QUEIROZ, G. GeoBR: Intercâmbio Sintático e Semântico de Dados Espaciais.**Informática Pública**, v.9, n. 2, p. 251–281, 2002.

LUDWIG, C.; ZIPF, A. Exploring regional differences in the representation of urban green spaces in OpenStreetMap, 2019. **Proceedings of the Geographical and Cultural Aspects of Geo-Information: Issues and SolutionsAGILE 2019 Workshop**, Limassol, Cyprus, 2019. p.1-3. Available in: https://archiv.ub.uni-heidelberg.de/volltextserver/27433/1/GeoCultGIS_Proceedings_2019.pdf.Accessed on: Feb 13, 2024.

MOONEY, P.; CORCORAN, P. Can Volunteered Geographic Information be a Participant in eEnvironment and SDI? In: Hřebíček, J., Schimak, G., Denzer, R. (eds) **ISESS 2011: Environmental Software Systems. Frameworks of eEnvironment. IFIP Advances in Information and Communication Technology book series (IFIPAICT)**, Berlin, Heidelberg, 2011. v.359, p. 115-1221. DOI: https://doi.org/10.1007/978-3-642-22285-6_13.

NEUMAIER, S.; SAVENKOV, V.; POLLERES, A. Geo-Semantic Labelling of Open Data. (null Procedia Computer Science, Org.)**Proceedings of the 14th International Conference on Semantic Systems (SEMANTiCS)**,Vienna, Austria, 2018. p. 9–20. Procedia Computer Science, vol. 137. Available in: http://www.sciencedirect.com/science/journal/18770509/137. Accessed on: May 04, 2024.

NOVACK, T.; GRINBERGER, A. Y.; SCHULTZ, M.; ZIPF, A.; MOONEY, P. The Geographical and Cultural Aspects of Geo-Information: **Proceedings of the Geographical and Cultural Aspects of Geo-Information: Issues and Solutions AGILE 2019 Workshop**, Limassol, Cyprus, 2019.p.10–13. Available in: https://agile-gi.eu/conferences/proceedings/proceedings-2019. Accessed on: May 04, 2024.

NOVACK, T.; PETERS, R.; ZIPF, A. Graph-Based Matching of Points-of-Interest from Collaborative Geo-Datasets. **ISPRS International Journal of Geo-Information**, v. 7, n. 3, p. 117, 2018.

OPENGIS®. **Implementation Standard for Geographic information – Simple feature access – Part 1: Common architecture**, 2011. Herring, J. Available in: https://www.ogc.org/standard/sfa/. Accessed on: Mar 14, 2024.

OPENGIS®. **Geographic information – Features and geometry – Part 1: Feature models**, 2018. Herring, J. Available in: https://docs.ogc.org/as/17-087r13/17-087r13.html. Accessed on: Mar 14, 2024.

OpenStreetMap Wiki. Available in: https://wiki.openstreetmap.org/wiki/Main_Page. Accessed on: Feb 02, 2024.

INDE. Infraestrutura Nacional de Dados Espaciais. **Portal INDE**.Available in: https://inde.gov.br/. Accessed on: Feb 13, 2024.

SALVUCCI, G.; SALVATI, L. Official Statistics, Building Censuses, and OpenStreetMap Completeness in Italy. **ISPRS International Journal of Geo-Information**, v. 11, n. 1, p. 29, 2022. Multidisciplinary Digital Publishing Institute.

SCHUMACHER, U. The Urban Mask Layer as Reference Geometry for Spatial Planning: Moving from

German to European Geodata. **KN - Journal of Cartography and Geographic Information**, v. 71, n. 2, p. 83–95, 2021.

SIEBER, R. E.; JOHNSON, P. A. Situating the Adoption of VGI by Government. In: Sui, D. Z.; Elwood, S.; Goodchild, M.**Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice**, Dordrecht, Springer, 2013. DOI: https://doi.org/10.1007/978-94-007-4587-2_5.

TRAJCEVSKI, G.; BALASUBRAMANI, B. S.; CRUZ, I. F.; TAMASSIA, R.; TENG, X. Semantically Augmented Range Queries over Heterogeneous Geospatial Data. Proceedings of the 28th International Conference on Advances in Geographic Information Systems. **Anais.**, SIGSPATIAL ’20. New York, NY, USA, 2020. Association for Computing Machinery. p.68–77 DOI: https://doi.org/10.1145/3397536.3422271.

TSIAVOS, P.; NEDAS, K.; PAIDIADITI, K.; ATHANASIOU, S. Cultivating Open Geodata Ecologies: Lessons from the implementation of Directive 2007/2/EC. **Geographic Data and the Law. Defining New Challenges**, 2012. Leuven University Press Belgium.

TVERSKY, B. & HEMENWAY, K. Objects, Parts, and Categories. **Journal of Experimental Psychology: General**, v. 113, p. 169–192, 1984.

UFPR. Universidade Federal do Paraná. **Base de Dados Científicos (BDC) da UFPR**. Available in: <https://bdc.c3sl.ufpr.br/discover?query=Adriana+Alexandria+Machado&scope=123456789/3&filtertype=author&filter_relational_operator=equals&filter=MACHADO%2C+ADRIANA+ALEXANDRIA>. Accessed on: Jan 04, 2023.

VILCHES-BLÁZQUEZ, L. M.; RAMOS, J. Á. Semantic conflation in GIScience: a systematic review. **Cartography and Geographic Information Science**, v. 48, n. 6, p. 512–529, 2021. Taylor & Francis. DOI: https://doi.org/10.1080/15230406.2021.1952109.

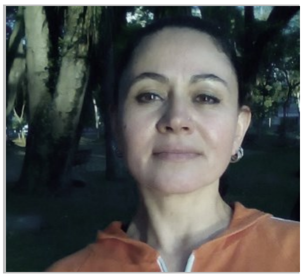
WACHE, H.; VÖGELE, T.; VISSER, U.; et al. Ontology-Based Integration of Information - A Survey of Existing Approaches. In: Gómez Pérez, A.; Gruninger, M.; Stuckenschmidt, H.; Uschold, M.(eds.), **Proceedings of the IJCAI-01 Workshop onOntologies and Information Sharing**, Seattle,USA, 2001. p. 108–119. Avaiable in: https://dblp.org/rec/conf/ijcai/2001ois.html. Accessed on: May 04, 2024.

XAVIER, E. M. A.; ARIZA-LÓPEZ, F. J.; UREÑA-CÁMARA, M. A. A Survey of Measures and Methods for Matching Geospatial Vector Datasets. **ACM Computing Surveys**, v. 49, n. 2, p. 1–34, 2017.

ZHANG, C.; ZHAO, T.; LI, W. The framework of a geospatial semantic web-based spatial decision support system for Digital Earth. **International Journal of Digital Earth**, v. 3, n. 2, p. 111–134, 2010. Taylor & Francis. DOI: https://doi.org/10.1080/17538940903373803.

ZHANG, H.; MALCZEWSKI, J. Accuracy Evaluation of the Canadian OpenStreetMap Road Networks. **International Journal of Geospatial and Environmental Research**, v. 5, n. 2, p. 1–14, 2017.

Main autor biography



Adriana Alexandria Machado, was born in Curitiba, Paraná, Brazil in January 03th, 1973. Cartographer and Surveyor Engineer graduated from the Federal University of Paraná (UFPR) and PhD in Geodetic Sciences from the Postgraduate Program in Geodetic Sciences at UFPR. Experience in Geosciences with an emphasis on Cartography, VGI, Semantics and Spatial Data Infrastructure.