Article

Assessment and Visualization of OSM Consistency for European Cities

Dimitra Zacharopoulou , Andriani Skopeliti * and Byron Nakos

Cartography Laboratory, School of Rural and Surveying Engineering, National Technical University of Athens, 15780 Zografou, Greece; dimitrazacharopoulou@gmail.com (D.Z.); bnakos@central.ntua.gr (B.N.)

* Correspondence: askop@survey.ntua.gr; Tel.: +30-210-772-2639

Abstract: Volunteered Geographic Information (VGI) is a widely used data source in various fields and services, such as environmental monitoring, disaster and crisis management, SDI, and mapping. Quality is a critical factor for the usability of VGI. This study focuses on evaluating logical consistency based on the topological relationships between geographic features while considering semantics. It addresses internal (i.e., between thematic layers) and external (i.e., between specific features from different thematic layers) logical consistency. Attribute completeness is computed to support the use of semantics. A tool for assessing the consistency and attribute completeness is designed and implemented in the ArcGIS environment. An open-source web mapping application informs users about VGI consistency with multiscale visualization and indices. Data from OpenStreetMap (OSM), one of the most popular collaborative projects, are evaluated for six European cities: Athens, Berlin, Paris, Utrecht, Vienna, and Zurich. The case study uses OSM-derived data, downloaded from Geofabrik and organized into thematic layers. OSM's consistency is evaluated and visualized at the regional, city, and feature levels. The results are discussed and conclusions on attribute completeness and consistency are derived.

Keywords: VGI quality; consistency; visualization; POI; volunteered geographic information; crowdsourcing; data quality; intrinsic quality assessment; contributors; OpenStreetMap

Citation: Zacharopoulou, D.; Skopeliti, A.; Nakos, B. Assessment and Visualization of OSM Consistency for European Cities. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 361. <https://doi.org/10.3390/ijgi10060361>

Academic Editors: Gavin McArdle, Bianca Schoen-Phelan and Wolfgang Kainz

Received: 11 April 2021

Accepted: 20 May 2021

Published: 25 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The ease of geolocation through mobile phones and the ability to create and share online maps and georeferenced photos have resulted in a phenomenon known as Volunteered Geographic Information (VGI) [1]. VGI encompasses a plethora of crowdsourced geographic data, such as geo-tagged photos from Panoramio and Flickr, spatial data and maps from OpenStreetMap (OSM) and Wikimapia, and 3D data in OSM-3D and OSM2World. VGI is a widely used data source in various fields and services, such as environmental monitoring [2], disaster and crisis management [3], SDI [4], mapping [5], and recreational activities management [6]. As with any dataset, the quality and the needs of the user determine the suitability for use.

When using VGI maps, the user rarely questions the reliability of the locations or the attributes because spatial data are considered reliable due to the high-quality visualization. However, the average map reader can easily spot consistency issues. In OSM, the most popular VGI platform, several inconsistency issues can be observed. For example, a bus stop may be located on the street axis (Figure 1a), and a shoe shop may be located outside the building (Figure 1b). Therefore, consistency evaluation becomes a priority in VGI quality assessment. In the pre-VGI era, many research projects [7–10] have shown that quality visualization supports decision-making and leads to significantly better decisions. Therefore, in the VGI case, where quality is often questionable, the visualization of VGI's consistency is considered as necessary as the assessment per se.

Figure 1. Consistency errors in OSM in Vienna. A bus stop appears inside a road (a), whereas, in reality, it is on the pedestrian street (<https://goo.gl/maps/DjltvmhuHwgR3Lvy6> accessed on 10 May 2021) (b). A shoe shop appears outside of the building (c), whereas, in reality, it is inside (<https://goo.gl/maps/fbnULEMyZ4MJAgbq5> accessed on 10 May 2021) (d).

The purpose of this paper is to evaluate VGI's quality, focusing on the consistency quality parameter. Intrinsic measures and indicators, which do not require comparison with authoritative data, are adopted. Data-driven indicators and measures are formed based on the topo-semantic relationships between features. A custom tool designed and developed in ArcGIS assesses violations of topo-semantic consistency and evaluates attribute completeness. In addition, a web mapping application informs the user about OSM consistency with tables and multiscale visualization. In a case study, OSM data consistency is investigated for six European cities: Athens, Berlin, Paris, Utrecht, Vienna, and Zurich.

Of particular interest is the quality assessment for points of interest (POIs). A POI is a specific point that may be helpful or interesting (e.g., a monument, a restaurant, a supermarket, a pharmacy, or a school). POIs are valuable resources utilized in daily life (e.g., navigation, social networks, and tourism), as well as in various commercial domains (e.g., logistics, advertising, or geomarketing) [11]. Research on Public Participation Geographic Information Systems (PPGIS) has shown that from a cognitive aspect, points are easier for users to understand than polygons or lines [12]. Among the open POI data sources, OSM is the most prominent [13]. In OSM, POIs are those that are mainly provided by users and updated more frequently. Quality evaluation of OSM POIs is of great importance and has been a research topic in the past [14–19].

This paper is organized as follows: Section 2 provides additional background material on VGI quality, quality visualization, and topo-semantic consistency assessment; Section 3 describes the topo-semantic consistency indicators for OSM, the custom developed GIS tool, and the design of a web map application for quality visualization; Section 4 describes

the case study results in six European cities and the web map application; and Section 5 discusses the results and presents future plans.

2. Background

2.1. VGI Quality Assessment

While research on VGI quality assessment matures, it is widely accepted that additional data quality indicators are required to complement those proposed in the ISO framework (e.g., ISO 19157). In many situations, comparison with authoritative datasets is impossible [20]. In contrast, the characteristics and nature of VGI enable the use of indicators that do not usually make sense when applied to data created by professionals and permit the development of automated approaches. Several researchers have classified the proposed indicators into groups [20] according to their nature [21–25]. Data-based indicators appear in categories labeled geographic consistency [21], internal quality indicators [22], internal quality [23], data indicators [24], and indicators of quality [25]. Data-driven indicators are additionally elaborated and classified into groups based on those in [20]:

- Coherence with other sources of corresponding data (which are not considered as references) through comparison (e.g., geometric attributes such as the distance between corresponding elements or overlaps);
- External logical consistency between VGI and non-corresponding data available in other data sources;
- Internal logical consistency of the VGI dataset itself;
- Metadata (e.g., the number of versions, features corrections, stability against changes, observation methods, used equipment, and date of observation).

This work pertains to indicators resulting from the VGI's internal logical consistency.

2.2. VGI Quality Visualization

Visualization can contribute to VGI usage. Appropriate visualization methods can support the interaction of the users involved [26]. For example, VGI visualization can strongly influence cognitive processing, such as memory performance [27]. In this framework, visualization is considered the key to communicating VGI quality to the crowd. Visualization can transform the quality from an ignored and complex data issue to an observable and vivid one. VGI data quality visualization can be leveraged as an awareness tool for the novice user, attracting the crowd's attention, communicating the quality, and stimulating improvements [28]. In addition, it provides an exploration tool that can help researchers study the appropriateness and the ability of indicators to represent quality, discover dependencies on extrinsic socioeconomic or demographic factors, and explore the spatial distribution and heterogeneity of VGI quality.

A thorough literature review [28] led to specific guidelines for VGI quality visualization in terms of user experience. In the present study, quality visualization is utilized as an awareness tool, and therefore intrinsic methods are recommended. However, extrinsic methods are sometimes preferred when a change in the original map symbolization is not desired, as in the OSM case. Extrinsic methods introduce new objects that work independently of the existing symbols to depict quality (e.g., abstract graphical objects such as circles or squares). According to established cartographic practices, the type of quality measure or indicator and the level of measurement (nominal, ordinal, interval, and ratio) play a major role in the selection of visual variables. In terms of user performance, color hue, value, and transparency provide the best alternatives. Coincident maps are preferable because integrating reliability into the display makes retrieval easier [29], especially for novice users. Advanced complexity can be abated with good cartographic design and interactivity (e.g., using on and off buttons) [28]. Finally, map scale plays an important role in the selection of the visualization method. Intrinsic methods are best suited for larger feature-level scales, while extrinsic methods, such as error grids, are preferable for global quality visualization in medium- and small-scale maps [30].

2.3. VGI Consistency

Inconsistency problems (e.g., sections of a network that are not connected, boundaries of a building and a city block that intersect, or land use polygons that overlap) are frequently observed in VGI. They have been studied in the past [16–18,31–35]. Inconsistency is sometimes related to the experience of the contributor [36]. The map scale (i.e., zoom level) also plays an important role. Some of these errors are detected and corrected during collection, as several editors perform validation that fixes invalid data (e.g., iD Editor and JOSM) [37].

Intrinsic consistency assessment can be based on the geographic properties of features [21] as expressed by spatial relationships [17,35] and topo-semantic relationships [16,18]. The first approach is cartography-oriented, as the reification of spatial relations follows the definition of spatial relations instances to guide generalization [38]. In [39], it is stated that the identification of implicit spatial relations helps to define good specifications for VGI contributors. In [40], the addition of integrity constraints to check on spatial relations in VGI contributing software (e.g., verify that roads do not cross buildings or bus stops are located on roads) was proposed. In [17], the proposed method used spatial integrity relations and automatically processed features to identify instances of these relations between them.

The second approach is geographic database-oriented and concerns topo-semantic consistency, as introduced in [41], for vector datasets. Topo-semantic consistency is considered a subset of logical consistency [42], but it concerns the correctness of the topological relationship between two objects according to their semantics. A building inside another building is undoubtedly an error, whereas a building inside a parcel is not, although in both cases, the relationship is a polygon inside a polygon [41]. Object semantics are mandatory to achieve the correctness of each relationship. Topological integrity constraints define rules based on data semantics. These constraints are adapted to the dataset under processing and can be also adapted to VGI. They can describe the consistency of geographic objects with other geographic objects of the same theme (intra-theme consistency) or geographic objects of different themes (inter-theme consistency). Check methods encounter two difficulties: the problem of exhaustiveness of all types of errors that are likely to occur and the problem of proving the completeness of each check method [41]. Several measures have been proposed in the literature to evaluate the degree of violation of topo-semantic constraints [43–47]. Since OSM can be considered both a map and a geographic database, it is noteworthy that the two approaches take into account the dual nature of the project.

Several constraints for the data arise from both approaches: two parcels must not overlap, a house must be inside a parcel, two roads must not be equal, a river and a road must not intersect except in the geographic object bridge, and land parcels must not overlap. Some studies have applied these constraints to evaluate consistency at the city level [16–18]. The topo-semantic consistency of OSM was evaluated in Paris [16,18] with specific constraints that checked POIs with buildings, roads, and railways. The results were reported with error percentages and proved that most of the tested features satisfied the constraints. In [17], spatial relations were applied for specific POIs such as amenities, ATMs, schools, and bus stops. The evaluation was enriched with the computation of the distance, which quantified the severity of the violation, and the examination of POIs in the same building. Satisfactory results for consistency were also extracted in this work. Although topo-semantic consistency has been successfully evaluated in these projects, there is a lack of communication of the spatial nature of the phenomenon to the user. This can be achieved through visualization.

In this paper, the term topo-semantic consistency check is adopted. Checks for different OSM entities (e.g., shops, bus stops, train stations, and buildings) are proposed. They are applied to six cities across Europe, and the results are reported and visualized in a web map application. Consistency is evaluated and visualized at the feature level (e.g., POI), city level (e.g., Athens), and regional level (e.g., Europe).

3. Assessment and Visualization of Topo-Semantic Consistency in OSM

OSM is a collaborative project that aims to create a free, editable map of the world. It is the largest and most diverse, comprehensive and up-to-date open-access geographic database. It is available with an open license that allows anyone to use, distribute, share, and customize the database, as long as the program and its community are mentioned [48]. In this research, OpenStreetMap was selected for VGI quality evaluation due to its popularity and richness of entities. Many web services provide OSM data for an area of the user's choice. OSM data can be extracted in the OSM data schema in XML from the official OSM site [49] or in a derived form from other web pages such as Geofabrik [50] or BBBike [51]. OSM-derived data is organized into thematic layers [52] and encoded in the shapefile format. Data organized into thematic layers are closer to the traditional data model for cartography or spatial analysis. However, it should be noted that derived OSM data may have several differences in geometry or attributes from the original data. On the other hand, less experienced users who use the derived OSM data are usually less familiar with the issue of VGI quality. It is really important to inform them. In this paper, the case study data were downloaded from GeoFabrik [50]. Data were organized into layers with points (POIs and places), polygons (buildings, land use, and nature) and lines (roads, railways, and waterways). POIs and places were considered as one layer in this work.

Consistency checks address OSM data quality and are considered as internal logical consistency checks and external logical consistency checks.

3.1. Internal Logical Consistency

Geographic data, to be considered reliable, must have correct topological relationships. Internal logical consistency refers to the topological relationships between the spatial entities in a thematic layer or between thematic layers. The semantic character of each layer is taken into account (i.e., the topologic relationship applies to all buildings without considering the building type). Specifically, for OSM data, the following topological constraints are applied [16]:

- Buildings must not overlap;
- Roads must not overlap, must not self-overlap, must not self-intersect, must not overlap with railways, and must not overlap with waterways.

In each case, the incorrect features are identified and flagged in the database. The results of the topological constraints are summarized for each city with error percentages. Regarding the overlap of polygon entities (i.e., buildings), the assessment of sliver polygons complements the violation of the topological constraints. Sliver polygons are small, narrow polygons that occur along the edges of overlapping polygons and cause topology errors. They occur due to positional errors or inappropriate zoom levels during the data collection and conversion of linear objects to polygons. The mathematical condition for the detection of sliver polygons is the ratio of the polygon perimeter to the square root of the polygon area [53]. The percentage of sliver polygons is also reported.

It must be noted that the buildings overlay constraint is in contrast to the OSM data model, which allows polygon overlay. However, for most mapping and spatial analysis projects, polygon overlay in some thematic layers (e.g., buildings) can cause data issues, and it is helpful to inform the user. Because derived data are used, these overlaps may not exist in the original OSM data.

3.2. External Logical Consistency

External logical consistency concerns the study of the validity of a spatial entity by combining the topological relationships and the semantics (topo-semantic relations). The plethora of tags in OSM provides a semantically rich dataset [54] and allows for investigating the topo-semantic relationships. The semantic information provided by the "Type" tag is elaborated with the geometric position. As a result, the criteria for assessing consistency are established. Specific checks evaluate the topology of spatial entities, taking into account data semantics as described by their attributes. Although checking attribute

completeness was not a specific task for this research, it should precede the use of tag values. Therefore, attribute completeness is calculated in terms of the percentage of data with “Type” tag values.

Many topo-semantic consistency checks can be formulated for POIs. They are based on semantic information derived from the “Type” tag and the spatial relationship with the semantically related layers. Thus, POIs are checked against specific thematic layers (e.g., buildings, nature, roads, or railways) [16]. For example, a cafeteria should be located inside a building and, at the same time, off the road, and a tram stop should be located on a railway line and, at the same time, outside buildings. To test the topo-semantic consistency, POIs are classified into six categories (Figure 2) (see Appendix A, Table A1):

- POIs must be inside buildings (Check 1: POIs must be inside buildings) (Figure 2a). This category includes POIs such as cafes, schools, pharmacies, and supermarkets, which should be located within the building polygons;
- POIs that are semantically related to the road network and must be outside the road network (Check 2: POIs must be outside of roads) (Figure 2b). This includes POIs such as bus stops, parking, and street lamps, which are related to the roads and are usually located very close to them but not on them;
- POIs that are semantically related to the road network and must be located outside buildings (Check 3: POIs must be outside of buildings) (Figure 2c). This includes POIs such as bus stops, junctions, and traffic lights, which should not be located within buildings;
- POIs that are semantically related to the road network and must be on the road network (Check 4: POIs must be on roads) (Figure 2d). This includes POIs such as junctions, traffic lights, and turning points that should be located on the road network;
- POIs must be outside the polygons of nature (Check 5: POIs must be outside of nature) (Figure 2e). This includes POIs such as department stores, hotels, car rentals, and cinemas, which must be outside the nature polygons;
- POIs that are semantically related to the rail network and must be located on the rail network (Check 6: POIs must be on railways) (Figure 2f). This includes only one kind of POI: the railway stops that should be located on railways.

After studying the values in the “Type” tag for POIs in some European cities, it was decided which ones would participate in each constraint. The formulation of the categories was based on common sense (i.e., what is usually expected) and the entity definition in the OSM wiki. In each case, the incorrect POIs were detected and marked in the database (Figure 2). Thus, topo-semantic consistency was assessed at the feature level. Therefore, errors could be visualized at the feature level in the context of quality communication. Some of the errors were exceptions (e.g., a bus stop in the central station building (Figure 2c). Others could be corrected by user choice with other VGI sources (e.g., Flickr geotagged photos) [16], satellite imagery, or ground truth.

Besides that, the distance of each POI to the thematic layer participating in the constraint was calculated [11]. Errors with small distances were due to the zoom level used during data entry [55]. The distance values supported the critique of the severity of the deviation at the map scale. Since the deviation was often minimal, it could be considered negligible at a map scale if the data were intended for portrayal only. Additionally, the distances could be compared to characteristic quantities at the city scale, such as the pavement size and road size [17]. For the use of the data in spatial analysis, of course, all deviations were errors.

The results of each topo-semantic check (Figure 3a) were summarized for each city with error percentages. In this way, the cities could be examined in comparison. The error and correct percentages could be visualized for a set of cities in a map, thus providing visualization at a regional level (e.g., Europe). However, they could not be summarized into a global Europe index because the cities were not systematically selected. In order to summarize the results of all topo-semantic checks and capture the spatial distribution of

inconsistency at the city level, a topo-semantic consistency error grid was proposed. Two types of topo-semantic consistency error grids were proposed. In the first one, called a “Consistency Grid (error existence)” (Figure 3b), the grid cell was marked as an error if an error existed in it. Thus, topo-semantic error existence was emphasized. In the second one, which was called “Consistency Grid (most frequent)” (Figure 3c), the grid cell was marked according to the most frequent value (error or correct). Thus, a more concise dataset was created. Finally, the topo-semantic consistency error grid allowed the calculation of a global error percentage for each city that expressed topo-semantic inconsistency based on all topo-semantic checks (Figure 3a).

Figure 2. Topo-semantic checks. Check 1: “POIs must be inside buildings” (a); Check 2: “POIs must be outside of roads” (b); Check 3: “POIs must be outside of buildings” (c); Check 4: “POIs must be on roads” (d); Check 5: “POIs must be outside of nature” (e); Check 6: “POIs must be on railways” (f).

Figure 3. The topo-semantic consistency error grids are calculated from POI errors (Zurich) (a), as is “Consistency Grid (error existence)” (b). If an error exists in the grid cell, the grid cell is marked as an error. In “Consistency Grid (most frequent)” (c), the grid cell is marked according to the most frequent value (error or correct).

3.3. A Tool for Checking Topo-Semantic Consistency

The “OSM consistency and completeness check” toolbox was designed and developed in ArcGIS 10.5 by utilizing the available geoprocessing tools and Model Builder [56,57]. It contained a set of checks (Figure 4) that evaluated OSM quality with indicators of completeness, consistency, and topological relationships, as explained earlier. The tool defined the rules applied to the features and exported the errors. Moreover, it reported statistics on quality evaluation for the area of interest. The “OSM consistency and completeness check” toolbox is available on Github (<https://cutt.ly/hbSqFkL> accessed on 21 May 2021) with an MIT license.

- ☐ Consistency and Completeness checks
 - ☐ Completeness check
 - Type field completeness
 - ☐ General Topology Checks (Internal)
 - ☐ Topology check for Network data
 - Topology check for Linear features(roads, railways, waterways)
 - ☐ Topology check for Polygon features
 - Topology check for Buildings
 - Topology check for Landuse
 - Topology check for Nature
 - ☐ Prepare data(optional)
 - 1.0 Create GDB with Dataset (optional)
 - 1.1 Rename FID to OBJECTID
 - 1.2 Different field
 - 2.Add new field with name "type" (optional)
 - 3.Merge all Point Features into One (optional)
 - 4.Select the area of interest (optional)
 - ☐ Topo - Semantic checks for POIs
 - 1.Add checks to POIs
 - 2.POIs VS Buildings, Nature and Networks

Figure 4. Consistency and completeness tool.

3.4. Design of a Web Mapping Application for the Visualization of the Topo-Semantic Consistency

A web mapping application was designed for the presentation and visualization of the topo-semantic consistency in the selected European cities. It provides the user with a better understanding of the VGI quality through visualization and interactivity. The user can be informed about the quantitative values and the spatial distribution of the inconsistency. The application offers functional visualization and information on data consistency. It has the following characteristics:

- Basemap: A list of the available basemaps (e.g., OSM) and the ground truth (e.g., satellite imagery) are provided to the user;
- Thematic maps: Quality visualization is overlaid on the basemap, and extrinsic techniques introduce new graphic objects. The use of OSM map tiles as a basemap does not allow intrinsic visualization methods. Quality visualization appears upon user request, and the OSM experience is not altered;
- Scale: Multiscale consistency visualization at the regional (i.e., Europe), city (e.g., Paris), and feature levels (e.g., POI) is provided. A scale bar in meters informs the user about the map scale;
- Legend: A visual explanation of the symbols used in the map is provided;
- Retrieval of consistency information at the city and feature levels;
- Interactivity and map navigation: An interactive graphic user interface and essential map navigation tools, such as zoom and pan, are available.

The application is map-centric. At the regional level, a map of Europe visualizes the results for each check and city. The cities are portrayed as points on this small-scale map (Figure 5a), and the error and correct percentages are portrayed. This map provides an overall picture of each check and allows comparison between European cities. When the user zooms in, inconsistencies are portrayed on the medium- and large-scale maps at the city level (Figure 5b) with an error grid and at the feature level (e.g., POIs) (Figure 5c) by symbolizing the erroneous ones.

Figure 5. Consistency visualization at different levels and scales: (a) regional (i.e., Europe), (b) city (e.g., Athens), and (c) feature levels (e.g., POI).

4. Case Study and Results

4.1. Data

Six European cities—Athens, Berlin, Paris, Utrecht, Vienna, and Zurich—were selected for the case study. The OSM data were downloaded (October 2019) from the Geofabrik website [50] in the shapefile format. The extracted data (Table 1) were organized into thematic layers as previously described. Table 1 reports the area examined and the number of features in each thematic layer. The downloaded OSM data used geographic coordinates in WGS 84, but a Cartesian coordinate system was assumed for the topo-semantic checks and the distance calculation. For each city, the geodetic reference system of the respective country was applied (Table 2).

Table 1. Area covered for each city, number of features in each thematic layer, and percentage of POIs checked.

City	Area (km ²)	POIs	Buildings	Roads	Nature	Rails	POIs Checked (%)
Athens	17.5	8428	33,939	5863	1	122	80
Berlin	107.6	101,560	62,597	45,167	4	3295	32
Paris	105	188,466	103,946	40,724	4	3185	34
Utrecht	285.6	18,969	244,301	44,331	6	1501	45
Vienna	22.5	34,438	16,671	12,182	2	1027	31
Zurich	112	31,668	53,339	37,243	6	3289	48

Table 2. Percentage of omission for buildings in “Type” and geodetic reference system for each city.

City	Percentage of Omission in “Type” for Buildings	Geodetic Reference System
Athens	96	GGRS87
Berlin	42	ETRS89/LCC Germany (E-N)
Paris	94	RGF93/Lambert-93
Utrecht	42	Amersfoort/RD New
Vienna	90	MGI/Austria Lambert
Zurich	72	CH1903/LV03

4.2. VGI Quality

4.2.1. Completeness Check and Information about the Type

The application of topo-semantic constraints required checking attribute completeness in advance. Therefore, the attribute completeness check informed the user about the omission in the “Type” tag for each thematic layer and the variety of values. In almost all thematic layers, the “Type” tag was highly populated with rich content. Only the “Buildings” thematic layer exhibited a significant lack of information (Table 2). The percentages of omissions were high, with higher rates in Paris (94%), Athens (96%), Vienna (90%), and Zurich (72%). For Berlin and Utrecht, the percentages were lower than 50%, indicating a lead in attribute completeness for these cities.

There was a wide variety of unique values in the “Type” tag for each thematic layer and city (Table 3). The highest variability was observed in Paris and Berlin, and the lowest was in Athens. The variability in POI tags led to an adjustment of the topo-semantic checks to cover POIs appearing across cities. The POI consistency checks referred to 74 different types in six European cities (Appendix A, Table A1), covering 50% of the types that appeared across cities. As a result, for each city, a subset of the POIs occurring in the case study area participated in the topo-semantic tests (Table 1). For Athens, a larger percentage of POIs was examined, while for the other cities, less than 50% were examined. The differences in the percentages were due to the POI semantic variability and the spatial distribution in the European cities studied.

Table 3. Number of unique values in the “Type” tag for each thematic layer and city.

City	POIs	Buildings	Roads	Railways	Nature
Athens	120	42	18	5	1
Berlin	155	140	24	5	1
Paris	154	91	27	6	4
Utrecht	150	76	26	3	1
Vienna	141	58	23	4	1
Zurich	154	62	26	6	1

4.2.2. Internal Topology Checks

Internal topology checks were applied to the buildings and roads. The percentage of overlapping polygons for buildings (Table 4) was around 1% for all cities, with lower values for Zurich and Utrecht. Some of these overlaps could be considered sliver polygons (Figure 6a), especially for Athens and Utrecht (Table 4). Sliver polygons may have originated from errors in VGI data collection or may have been the result of the original OSM data’s transformation into layers. On the other hand, non-sliver overlaps represent cases where overlapping polygons represent reality within the OSM data model (e.g., in Berlin, many polygons were encoded as “roof” and overlapped with the building polygons (Figure 6b)). The lower percentage of sliver polygons in Berlin, Vienna, and Zurich showed a lead in data consistency for buildings. The percentages of overlapping polygons and sliver polygons should therefore not be taken as an indication of quality problems. They describe some data characteristics that the data user should check in the context of his or her project. In cartography, for example, overlapping polygons require special management of symbols. These findings were informative only, and they were not used in the calculation of the error grid. Internal topology checks for roads in all cities resulted in percentages of less than 0.01%, indicating that there were practically no internal topology errors.

Table 4. Errors in internal topology checks for buildings.

	Percentage of Overlapping Polygons	Percentage of Sliver Polygons
Athens	0.75	43
Berlin	0.80	18
Paris	0.84	23
Utrecht	0.55	36
Vienna	1.13	19
Zurich	0.44	14

Figure 6. Polygon overlaps may be sliver polygons due to errors in data collection or OSM data transformation to layers (a) (Utrecht) or represent reality within the OSM data model (e.g., in Berlin, many polygons were encoded as “roof” and overlapped with the building polygons) (b).

4.2.3. Topo-Semantic Consistency

Topo-semantic Checks 1–6 were applied to the selected POI types (see Table A1 in Appendix A) and the associated thematic layers (i.e., buildings, roads, rails, and nature). Table 5 shows the error percentages for each check. The following was observed:

- Check 1 (POIs must be inside buildings): Berlin exhibited the most errors (15.9%) while Utrecht (2.4%) exhibited the fewest. The cities of Athens, Zurich, and Vienna had an error percentage of about 5%, and that of Paris was less than 20%. In this check, POIs on the borders of buildings were encountered as errors. If they were considered correct (see Check 1 with on border in Table 5), the percentages for Paris and Berlin diminished, and the values for all cities became similar, ranging from 2% to 6%;
- Check 2 (POIs must be outside of roads): The highest error rate was observed in Athens (35.8%), and the lowest was in Paris (1.9%). The cities of Berlin and Utrecht exhibited error rates of less than 7%, while Vienna and Zurich had higher rates around 20–30%. This check refers to POIs such as bus stops, stops, and parking, which should not be on the road. All types of POIs may appear as error or correct, and no conclusion in terms of “Type” could be extracted. These types of POIs are not visible in the satellite imagery, and therefore, they are positioned according to the user’s personal knowledge. Additionally, they may have been imported from existing datasets where they needed to be on the road axis;
- Check 3 (POIs must be outside of buildings): The error rate was less than 1% for all cities and could be considered insignificant;
- Check 4 (POIs must be on roads): The error rate was less than 1% for all cities and could be considered insignificant;
- Check 5 (POIs must be outside of nature polygons): The lowest percentages were observed while some cities had no error at all. It is worth noting that the nature polygons had low coverage in all cities, which could justify the negligible error percentages;
- Check 6 (POIs must be on the rail network): This was the category with the highest percentages (70–100%), and this is discussed in more detail in the following paragraphs.

Table 5. Percentage of POIs checked and error percentages for each check and city.

City	Check 1	Check 1 with on Border	Check 2	Check 3	Check 4	Check 5	Check 6
Athens	5.4	4.9	35.8	0.8	0.0	0.02	87.5
Berlin	15.9	5.6	2.3	0.4	0.2	0	100
Paris	9.9	2.9	1.9	0.4	0.7	0.003	38.3
Utrecht	2.4	2.1	6.2	0.7	0.5	0	22.5
Vienna	5.3	3.7	19.3	0.4	0.4	0	100
Zurich	4.8	4.0	26.4	0.1	0.2	0	86.1

For a meaningful discussion of the Check 1 errors and severity judgment at scale, the distance to the building border was calculated for the erroneous POIs (Table 6). From Table 6, it can be seen that the minimum, the mean, the median, and the standard deviation were comparable for all cities. The values for Athens, Berlin, Paris, Utrecht, and Vienna were considered more homogeneous, while the values for Zurich were high. Regarding the maximum distance values, extreme values occurred for all cities, while outliers were observed for Berlin and Zurich. The extreme values proved the existence of gross errors. Possibly, they were due to the fact that some POIs, but not the corresponding buildings, existed in the dataset (Figure 4a). This proves that the coverage and completeness of the thematic layers may differ in the same city.

Table 6. Statistics for POI distances from the building border (Check 1).

	Min	Distances >0 m				Percentage of POIS			Distances ≤11 m	
		Mean	Median	St. Dev.	Max	<2 m	<5.5 m	<11 m	Mean	St. Dev.
Athens	0.001	5.1	2.7	6.8	47.2	58	71	89	4.3	2.6
Berlin	0.001	6.1	1.9	11.9	176.5	86	89	94	4.3	1.7
Paris	0.001	6.8	2.2	7.4	99.9	87	90	94	4.3	1.9
Utrecht	0.001	8	3	13.1	73.9	57	64	80	4.6	2.2
Vienna	0.001	7.3	2.4	12	73.2	68	76	86	4.3	2.1
Zurich	0.003	9.7	5.8	13.9	121.2	46	56	76	6.5	5.8

In order to further comment on the distance values and distinguish gross errors from misplacements, it was necessary to identify the critical values. At the city scale, critical values could be set based on the average sidewalk width (e.g., 2 m) [58] and the average street lane width (e.g., 3.5 m) [59]. A critical value was set at 5.5 m by adding the sidewalk width and a single-lane street. It was considered the “on the same side of the street (not crossing to the opposite side)” limit. Another critical value was set at 11 m. It was the sum of the sidewalk, a two-lane street, and the opposite sidewalk. It was considered the “in the wider street area” limit. For all cities, almost half of the POIs were in the sidewalk area (<2 m). A critical percentage of 56–90% was on the same side of the street (<5.5 m), and a substantial percentage from 76% to 94% was in the wider street area. Consequently, POIs with distances larger than 11 m were completely misplaced, or the associated building did not exist in the dataset. POIs with distances greater than 11 m were excluded, and the distance statistics were recalculated. Based on these new statistics, it can be seen that the average distances and the standard deviations were homogeneous and smaller. Most of the cities had an average distance value of about four meters and a standard deviation of about three meters. The POIs appeared in OSM at zoom levels of 17, 18, and 19, which corresponded to 4 K, 2 K, and 1 K scales [60]. Therefore, these errors were visible to the map user. Zurich still showed large values that differed from those of the other cities.

In general, POIs with extreme values represented cases where buildings had not been digitized (Figure 7a) or POIs were misplaced in the building area (e.g., into the patio or inner courtyard) (Figure 7b). Such cases can be identified using satellite imagery. Thus, POI errors can indicate areas with omitted buildings and support the update of the OSM dataset. Of course, there are always exceptions to the rules (e.g., restaurants in gardens that produce false errors (Figure 7c) or shops in railway stations (Figure 7d)).

For the discussion of the large error percentages for Check 6, the distances to railway lines were computed (Table 7). According to the OSM wiki, the project community defines a railway station as the place where trains stop and boarding and disembarking occurs. However, contributors have assigned this tag to station entrances, resulting to errors. In such cases, the distance calculation is of great help, as errors associated with large distance values are likely due to points incorrectly placed at the station entrances. A small percentage of them were situated in the sidewalk area (0–59%), more than half of the POIs were located on the same side of the street (<5.5 m), and more than 70% or even 100% were in the greater street area. Points with distances greater than 11 m were excluded, and the distance statistics were recalculated. From these new statistics, it can be seen that the average distances and the standard deviation were smaller than the original ones and more homogeneous. Moreover, the values of the distance statistics for the POIs in the wider street area were similar to those computed for Check 1. Thus, a four meter positional error on average could be considered for the POIs.

Figure 7. Errors for POIs in Check 1 resulting from (a) building omission (Athens), (b) POIs misplaced in the buildings (e.g., in the patio (Athens)), (c) false errors resulting from exceptions to the rules (e.g., restaurants or kiosks in parks (Paris)), (d) shops in railway stations (Zurich), and (e) bars in the riverfront (Paris).

Table 7. Statistics for POI distances from railways (Check 6).

	Distances (m)					Percentage of POIs			Distances ≤ 11 m	
	Min	Mean	Median	St. Dev.	Max	<2 m	<5.5 m	<11 m	Mean	St. Dev.
Athens	1.2	2.97	2.09	1.8	6.48	43	86	100	2.97	1.8
Berlin	0.4	4.2	4.56	1.9	9.88	20	71	80	4.21	1.9
Paris	0.13	4.8	1.75	7.60	36.90	59	78	89	2.51	1.5
Utrecht	2.09	6.9	5.01	5.22	20.79	0	56	89	5.16	1.9
Vienna	0.39	4.7	3.25	3.49	14.75	23	65	93	4	2.5
Zurich	0.84	7.3	4.97	5.59	20.22	19	58	74	4.35	2.4

In conclusion, small error percentages proved the validity of the topo-semantics checks formulation, because they attested that the constraints were compatible with the data. Sometimes, there were special cases due to the pluralism in each city that resulted in errors. From the error percentages (Table 5), Utrecht had the lowest number of errors. On the contrary, the city with the most significant number of errors was Athens. Apart from this general remark, there was no general trend in the indicators per city. A visualization that portrays the spatial distribution of errors will contribute to a complete picture of topo-semantic consistency.

Topo-semantic checks assessed the erroneous POIs, which could be visualized at large scales on the OSM. Thus, visualization at the feature level was achieved. Additionally, the error percentages (Table 5) summarize the check results at the city level. The error percentages for each city could be visualized on a small-scale map, where cities are por-

trayed as points with point symbols (e.g., pie charts). Thus, visualization at the region level was achieved. For the topo-semantic consistency portrayal at the city level (medium scale), topo-semantic consistency error grids were calculated for each city based on all erroneous POIs. The grids had a 111 m resolution (0.001 degrees at the equator) for the Web Mercator projection utilized for web maps. The Consistency Grid allowed for the calculation of a global topo-semantic error percentage for each city (Table 8). Utrecht had the best score based on the “Consistency Grid (error existence)”, and Paris had the best score for “Consistency Grid (most frequent)”. Athens had the worst score for both grids, and Utrecht had average percentages. However, a larger percentage of POIs was checked for Athens (80%), and a smaller one (45%) appeared for Utrecht. Berlin, Paris, and Utrecht exhibited average values in both indices and could be considered to be of acceptable consistency.

Table 8. Global topo-semantic error percentages.

City	Consistency Grid (Error Existence)	Consistency Grid (Most Frequent)
Athens	29.2	10.3
Berlin	10.3	3.2
Paris	10.0	2.1
Utrecht	6.8	4.8
Vienna	21.0	5.0
Zurich	14.54	6.0

4.3. Presentation of the Web Mapping Application for Quality Visualization

The OSM quality visualization web mapping application can be accessed at <https://cutt.ly/PbQuARB> (accessed on 21 May 2021). The main menu of the web app appears on the left (left side menu). It contains seven tabs indicating Europe and the European cities studied, as well as the “About” table. The application is map-centric, and the map occupies most of the application window. The layer control appears in the upper right corner of the map. The user can choose a basemap between OpenStreetMap, OpenStreetMap B&W, and ESRI World Imagery.

The map of Europe is the first map of the web application. It portrays the statistics for each check (i.e., the percentage of error at the region level) with pie charts. The size of the pie chart is analogous to the number of POIs in each city, and the sizes of the sectors are related to the percentage of the erroneous and correct POIs. The user is informed about the severity of the inconsistencies and the number of checked points (Figure 8). The results can be compared across cities, and the user gets an overall picture. Additional information is displayed by clicking on the pie chart symbol (Figure 8). When the user selects a topo-semantic check from the layer control, the relevant thematic map is displayed (Figures A1 and A2 (Appendix B)). The legend appears below the left side menu.

When the user selects one of the cities from the left side menu, the map zooms to the extent of the city. At the city level, the topo-semantic consistency is portrayed with the “Consistency Grid (error existence)” or “Consistency Grid (most frequent)” (Figures 9 and A3, Figures A4–A7). Green and red hues symbolize correct and error cells, respectively. Transparency in the grid’s portrayal permits seeing the OSM basemap. The user is informed about the error percentages of the grids from the legend.

Figure 8. At the regional level, topo-semantic errors for each check (e.g., “POIs must be inside buildings”) are portrayed with a pie chart for each city. Additional information is provided by clicking on the symbol.

Figure 9. “Consistency Grid (most frequent)” for Zurich, portrayed on the OpenStreetMap B&W basemap.

If the user zooms in and selects an entry from the layer list (e.g., “POIs must be inside buildings”), errors for this topo-semantic check are portrayed (Figure 10). For the symbolization of the topo-semantic check, the hue visual variable, which is suitable for data in the nominal scale, is selected. Circles with different hues are used as symbols. Through this symbol, the user can see the corresponding OSM symbol, which reveals the type of POI. The legend informs the user about the point symbols, the number of POIs, and the error percentages (Figures 10 and A8).

Figure 10. Errors from the topo-semantic consistency (Berlin) checks, symbolized with circles with different colors for each check as explained in the legend.

When a POI is clicked on, the user is informed about the “Name”, “OSM ID”, “Type”, “Check”, and distance to the features used in the topo-semantic consistency check (Figure 11a). Selection of World Imagery as a basemap may help the user understand the error source (e.g., misplacement or exception (Figures 11b and A9, Figures A10 and A11)).

Figure 11. When a POI is clicked on, the user is informed about the “Name”, “OSM ID”, “Type”, “Check”, and distance to the features used in the topo-semantic consistency check (a) (Paris). Selection of World Imagery as a basemap may help the user understand the error source (e.g., misplacement or exception) (b).

Consistency problems in the building polygons due to overlap can also be portrayed. From the layer list, the user can select “Buildings Overlaps” (Figure 12a) to see the common areas or “Overlapping Buildings” to see the original features that overlap (Figure 12b). Information about the buildings (e.g., “OSM ID” and “Type”) is available when the “Overlapping Buildings” option is active by clicking on it (Figure 12b).

Figure 12. Consistency errors in buildings (Athens) are portrayed with (a) “Buildings Overlaps” and (b) “Overlapping Buildings” (Athens).

The OSM data for buildings, roads, nature, and railways used in consistency checks can be portrayed (Figure 13a) on the map at the user’s choice (Figure 13b) to facilitate data understanding. In addition, the area of the city covered in the case study can be portrayed (Figures 13c, A10 and A11).

Figure 13. OSM data for buildings, roads, and railways (a) are portrayed at the user’s choice (b) to help data understanding. The area of the city covered in the case study can be portrayed as well (c) (Vienna).

The web application was built using both client-side and server-side tools for the publishing and portrayal of geographic data. Client-side tools are advantageous in terms of interactivity. The GeoJSON format for data encoding and the Leaflet library were utilized. Server-side tools are appropriate for large datasets and require a geographic data server (e.g., Geoserver). Regarding VGI quality symbolization, the Leaflet library was considered sufficient for GeoJSON symbolization. A Styled Layer Descriptor (SLD) was utilized for data published as a WMS service from Geoserver. The HTML and CSS languages were used to format and display the content of the web mapping application. JavaScript was used to enhance the functionality of the application. Additional screenshots of the web map application can be found in Appendix B.

5. Discussion, Conclusions, and Future Work

Although a variety of indicators and measures have been proposed in the past, the evaluation of VGI quality is still an open research topic. It is evident that visualization of VGI quality is of equal importance, as it works as an awareness tool for the novice user and an exploration tool for the expert. Consistency is an easily observed VGI quality element, and intrinsic assessment measures have been proposed. This research addresses consistency assessment based on topo-semantic constraints and visualization, along with an extensive case study.

5.1. Contribution to VGI Research and Limitations

In previous studies, topo-semantic constraints [16–18] have been studied for OSM data in Paris. In this study, topo-semantic constraints were adapted, extended, and applied to a more extensive dataset (i.e., six European cities). In previous works [16–18], error percentages were utilized to present topo-semantic check results, and conclusions were drawn for specific POIs. In this paper, topo-semantic checks were applied to a great variety of POIs. In addition to reporting on topo-semantic checks with error percentages, the topo-semantic consistency error grid was introduced. It summarized the results of all topo-semantic checks for an area and supported visualization at a regional level (i.e., a city). Further summarization was succeeded with the global error percentage extracted from the grid. A GIS tool that assesses attribute completeness and consistency based on topo-semantic checks for derived OSM data was developed in ArcGIS. A web mapping application that reported and visualized the VGI quality was implemented. Special attention has been given to POIs due to their importance, but other thematic layers were also evaluated. OSM consistency was analyzed for each city, and specific conclusions were extracted. The conclusions are compatible with those in [16–18]. Topo-semantic tests were enhanced with distance calculations [18] and evaluated with the help of critical values based on the sidewalk and street width. Three critical values were proposed: in the sidewalk (<2 m), on the same side of the street (<5.5 m), and in the wider street area (<11 m). An average position error for POIs was assessed after removing the outliers. Finally, a multiscale visualization of OSM consistency was presented. The results were portrayed on a small-scale map (at a regional level, such as Europe), where cities were considered points with pie charts that symbolized the correct or error percentages and the number of checked POIs. On a medium-scale map, where the city was regarded as an area (at city level), the topo-semantic consistency error grid was portrayed. Finally, at a large scale, the errors were displayed at the feature level (e.g., POI).

The research in this paper was based on two pillars: the GIS tool for consistency assessment and the web app for consistency visualization. Consistency assessment was implemented with the “Consistency and Completeness checks” tool developed in the ArcGIS environment. The tool allows the application of constraints in OSM data for any geographic area and the assessment of VGI quality with completeness and consistency indicators. The GIS tool has proven to be particularly useful. It can help users who may not have specific knowledge but need to be informed about the consistency of VGI data. Based on the completeness and logical consistency results, the user can decide on data suitability

in the framework of a project. The tool is compatible with derived OSM data downloaded from Geofabrik and does not require deep knowledge of ArcGIS or OSM. It can be applied to data from other sources (e.g., ground collected data), as long as the “Type” attribute and corresponding values are present. Based on the errors assessed by the tool, the user can correct the data based on ground truth or the model of his project. The “Consistency and Completeness checks” tool is available on Github. It is in an editable format (i.e., Model Builder Tool) and it can be modified; checks can be updated, and new checks can be added.

The contribution of the web map application is equally significant. The web map application presents with tables and thematic maps the topo-semantic consistency results. It offers to the user a better understanding of VGI quality through multiscale visualization and information retrieval. Consistency visualization is achieved at the regional (i.e., Europe), city (e.g., Paris), and feature levels (e.g., POI). Quality visualization is overlaid on the basemap, and the OSM experience is not altered. The web map app has the qualities of a modern web app such as interactivity, basemap selection, multiscale visualization, information retrieval, and essential map tools (e.g., scale bar, legend, and map navigation).

Two research decisions need to be discussed: the selection of the cities and the use of derived OSM data. In the case study, the assessment and the visualization of the VGI consistency were performed for six European cities: Athens, Berlin, Paris, Utrecht, Vienna, and Zurich. In order to test the validity of the proposed consistency assessment and visualization methods, a case study in a broad dataset covering several cities was deemed necessary. These cities were capitals or major cities. In the past, research on VGI quality assessment was conducted for the countries of these cities (e.g., Austria, Germany, Greece, France, Netherlands, and Switzerland) [24]. Finally, they can be considered representative of the European reality (north–south). The results of the cities were not combined into a European index; therefore, the distribution, the identity, and the specific characteristics (e.g., urban history, structure, and planning system) of the cities were not crucial. Other cities could have been selected without changing the validity of the study. Another issue is the use of derived OSM data from the Geofabrik site instead of the original data from the OSM site. As mentioned earlier, derived data were chosen because users prefer data organized in thematic layers for most projects. Additionally, since this is usually the choice of less-experienced users, it is important to inform them about the VGI quality by assessing and visualizing inconsistencies.

5.2. Conclusions

Several conclusions were drawn from the case study based on the assessment of the consistency for the six cities:

- Attribute completeness: This was present in most layers. In contrast, the percentages of omissions were high for buildings, but Berlin and Utrecht had percentages lower than 50%, signaling a lead in attribute completeness for these cities;
- Building overlap: Small percentages were found for all cities. Berlin, Vienna, and Zurich exhibited a lead in data consistency for buildings;
- Road overlap: Insignificant values were found for all cities;
- Regarding POIs that must be inside buildings, the error percentages were similar for all cities, ranging from 2% to 6%. Some errors were caused by the omission of buildings. It can be seen that for all cities, almost half of the POIs were located in the sidewalk area (<2 m), a critical percentage from 56% to 90% was on the same side of the street (<5.5 m), and an essential percentage from 76% to 94% was in the broader street area. Excluding outliers, an average position error was estimated at four meters;
- For POIs that must be outside of roads and POIs that must be on rails, significant errors were observed in all cities;
- Topo-semantic consistency error grids: Utrecht had the best score based on the “Consistency Grid (error existence)”, and Paris had the best score for “Consistency Grid (most frequent)”. Athens had the worst scores. Berlin, Paris, and Utrecht exhibited average values in both grids and could be considered as having acceptable consistency.

General remarks can be also concluded. Small error percentages prove the validity of the topo-semantics checks formulation, as they prove that the constraints are compatible with the data. They also prove the consistency of the OSM data. Exceptions to the rules (e.g., restaurants in gardens) may produce false errors. Extreme values in the distances calculated for POIs may originate from gross errors in the collection, POI misplacement in the building area (e.g., in the patio), and the omission of the corresponding buildings from the dataset. As a result, the coverage and completeness of the thematic layers (i.e., POIs and buildings) in the same city may differ. POI errors can be used to indicate areas with omitted buildings and support OSM data updates. The omission of buildings can be corrected in combination with the satellite imagery. Consequently, the assessment of topo-semantic errors can support the improvement of all thematic layers involved in them. Significant error percentages for “POIs that must be outside of roads” can be explained by the fact that these POIs are not visible on the satellite imagery, and thus they are positioned according to the user’s knowledge. Additionally, they may have been imported from existing datasets with entity definitions different from OSM. The high percentage of errors in railway stations was due to their malposition in the entrance of the stations rather than where the trains stopped and boarding and disembarking occurred. It seems that users mainly rely on world perception based on their experience and visual stimuli. For the average person, the station is where he or she can enter the station or where he or she sees the station sign. They probably do not read the OSM entity definition when they capture something so obvious to them. The random spatial distribution of POI topo-semantic errors advocates for their visualization on the map and the adequacy of the error grid.

5.3. Proposals

In the future, several improvements can be made. It would be advisable to use the original OSM data organized as entities, rather than the derived data in thematic layers. It would be interesting to review the differences with the present results. It was found that some of the errors assessed in this study no longer exist in OSM. This is because the corresponding features do not exist anymore or have been corrected, since OSM is continuously updated. The list of errors should be brought up to date with OSM by periodic evaluations. Otherwise, a web tool for on-the-fly assessment based on the original OSM data is the best solution. Another issue is the formulation of the topo-semantic constraints. They can be alternatively extracted from the data by analyzing the topological position of most POIs for each type. As the GIS tool is extensible, updating the topo-semantic checks is feasible. Rules enrichment can address more POI types, add new topo-semantic constraints, and address OSM objects portrayed at smaller scales or in areas with different characteristics (i.e., rural). Ideally, indices for other quality parameters should also be included. Enriching the tool would also enhance the web mapping application by visualizing errors for other quality elements. Application to a broader set of cities chosen in a more systematic way (e.g., all capitals, all major cities in Europe, or internationally) will result in a more comprehensive study of the map state. The results can also be used to study how users operate in different regions.

Author Contributions: Conceptualization, Andriani Skopeliti, Dimitra Zacharopoulou, and Byron Nakos; software, Dimitra Zacharopoulou; formal analysis, Dimitra Zacharopoulou and Andriani Skopeliti; writing—original draft preparation, Andriani Skopeliti; writing—review and editing, Dimitra Zacharopoulou, Andriani Skopeliti, and Byron Nakos; supervision, Andriani Skopeliti and Byron Nakos. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Topo-Semantic Tests for POIs

Table A1. Topo-semantic checks 1–6, which are applied to the POIs with specific “Type” values.

OSM Type	Check 1	Check 2	Check 3	Check 4	Check 5	Check 6
arts_centre	x					
atm	x					
bakery	x					
bank	x					
bar	x					
beauty_shop	x					
beverages	x					
bicycle_rental		x				
bicycle_shop	x					
bookshop	x					
bus_stop		x	x		x	
butcher	x				x	
cafe	x				x	
camera_surveillance		x				
car_dealership	x	x			x	
car_rental		x			x	
cinema	x				x	
clothes	x				x	
college	x				x	
community_centre	x				x	
computer_shop	x				x	
convenience	x				x	
crossing			x	x		
dentist	x				x	
department_store	x				x	
doctors	x				x	
doityourself	x				x	
fast_food	x				x	
fire_station	x				x	
florist	x				x	
fuel		x			x	
furniture_shop	x				x	
gift_shop	x				x	
greengrocer	x				x	
guesthouse	x				x	
hairdresser	x				x	
hostel	x				x	
hotel	x				x	
jeweller	x				x	
kindergarten	x				x	
kiosk					x	
laundry	x				x	
library	x				x	
mobile_phone_shop	x				x	
motorway_junction			x	x		
museum	x				x	
nightclub	x				x	
optician	x				x	
outdoor_shop	x				x	
parking		x			x	
parking_bicycle		x			x	
parking_underground		x				
pharmacy	x				x	
pitch					x	

Table A1. Cont.

OSM Type	Check 1	Check 2	Check 3	Check 4	Check 5	Check 6
police	x				x	
post_office	x				x	
pub	x				x	
railway_station						x
restaurant	x				x	
school	x				x	
shoe_shop	x				x	
sports_centre	x				x	
sports_shop	x				x	
stop		x	x			
street_lamp		x	x			
supermarket	x				x	
taxi			x		x	
theatre	x				x	
toy_shop	x				x	
traffic_signals			x	x		
travel_agent	x				x	
turning_circle			x	x		
vending_any					x	
veterinary	x				x	

Appendix B. The Web Mapping Application

Some characteristics are presented in screenshots from the web map application in the following figures.

Figure A1. At the regional level, topo-semantic errors for “POIs must be outside of roads” are portrayed with a pie chart for each city.

Figure A2. At the regional level, topo-semantic errors for “POIs must on Railways” are portrayed with a pie chart for each city.

Figure A3. Consistency Grid (error existence) for Athens.

Figure A6. Consistency Grid (error existence) for Utrecht.

Figure A7. Consistency Grid (error existence) for Vienna.

Figure A8. Errors from the topo-semantic consistency check “POIs must be inside of buildings” for Vienna. The “Buildings” layer is also portrayed.

Figure A9. Errors from the topo-semantic consistency check “POIs must be outside of buildings” for Paris are portrayed on the World Imagery basemap. The “Buildings” layer is portrayed as well. Additional information is provided by clicking on each error.

Figure A10. Errors from the topo-semantic consistency check “POIs must be on Railways” for Zurich. The “Railways” layer is portrayed as well. Additional information is provided by clicking on each error.

Figure A11. Errors from the topo-semantic consistency check “POIs must be outside of Roads” for Berlin are portrayed on the BW OSM basemap. The “Roads” layer is portrayed as well. Additional information is provided by clicking on each error.

References

1. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *Geo. J.* **2007**, *69*, 211–221. [\[CrossRef\]](#)
2. Connors, J.P.; Lei, S.; Kelly, M. Citizen science in the age of neogeography: Utilizing volunteered geographic information for environmental monitoring. *Ann. Assoc. Am. Geogr.* **2012**, *102*, 1267–1289. [\[CrossRef\]](#)
3. Klonner, C.; Marx, S.; Usón, T.; de Albuquerque, J.P.; Höfle, B. Volunteered geographic information in natural hazard analysis: A systematic literature review of current approaches with a focus on preparedness and mitigation. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 103. [\[CrossRef\]](#)

4. Olteanu-Raimond, A.-M.; Laakso, M.; Antoniou, V.; Fonte, C.C.; Fonseca, A.; Grus, M.; Harding, J.; Kellenberger, T.; Minghini, M.; Skopeliti, A. VGI in National Mapping Agencies: Experiences and Recommendations. In *Mapping and the Citizen Sensor*; Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C.C., Antoniou, V., Eds.; Ubiquity Press: London, UK, 2017; pp. 299–326. [\[CrossRef\]](#)
5. See, L.; Estima, J.; Pöddör, A.; Arsanjani, J.J.; Bayas, J.-C.L.; Vatseva, R. Sources of VGI for Mapping. In *Mapping and the Citizen Sensor*; Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C.C., Antoniou, V., Eds.; Ubiquity Press: London, UK, 2017; pp. 13–35. [\[CrossRef\]](#)
6. Santos, T.; Mendes, R.N.; Vasco, A. Recreational activities in urban parks: Spatial interactions among users. *J. Outdoor Recreat. Tour.* **2016**, *15*, 1–9. [\[CrossRef\]](#)
7. MacEachren, A.M.; Brewer, C.; Pickle, L.W. Mapping health statistics: Representing data reliability. In Proceedings of the 17th International Cartographic Conference, Barcelona, Spain, 3–9 September 1995; pp. 311–319.
8. Leitner, M.; Buttenfield, B.P. Guidelines for the display of attribute certainty. *Cartogr. Geogr. Inf. Sci.* **2000**, *27*, 3–14. [\[CrossRef\]](#)
9. Cliburn, D.C.; Feddema, J.J.; Miller, J.R.; Slocum, T.A. Design and evaluation of a decision support system in a water balance application. *Comput. Graph.* **2002**, *26*, 931–949. [\[CrossRef\]](#)
10. Deitrick, S.A. Uncertainty visualization and decision making: Does visualizing uncertain information change decisions? In Proceedings of the 23rd International Cartographic Conference, Moscow, Russia, 4–10 August 2007; pp. 4–10.
11. Patroumpas, K.D.; Skoutas, G.; Mandilaras, G.; Athanasiou, S. Exposing Points of Interest as Linked Geospatial Data. In Proceedings of the 16th International Symposium on Spatial and Temporal Databases, Vienna, Austria, 19–21 August 2019; pp. 21–30.
12. Brown, G.C.; Pullar, D.V. An evaluation of the use of points versus polygons in public participation geographic information systems using quasi-experimental design and Monte Carlo simulation. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 231–246. [\[CrossRef\]](#)
13. Patroumpas, K.; Georgomanolis, N.; Stratiotis, T.; Alexakis, M.; Athanasiou, S. Exposing INSPIRE on the Semantic Web. *J. Web Semant.* **2015**, *35*, 53–62. [\[CrossRef\]](#)
14. Mulligann, C.; Janowicz, K.; Ye, M.; Lee, W.C. Analyzing the spatial-semantic interaction of points of interest in volunteered geographic information. In Proceedings of the International Conference on Spatial Information Theory COSIT, Belfast, ME, USA, 12–16 September 2011; pp. 350–370.
15. Jonietz, D.; Zipf, A. Defining fitness-for-use for crowdsourced points of interest POI. *ISPRS Int. J. Geo-Inf.* **2016**, *59*, 149. [\[CrossRef\]](#)
16. Antoniou, V.; Skopeliti, A.; Fonte, C.C.; See, L.; Alvanides, S. Using OSM, geo-tagged Flickr photos and authoritative data: A quality perspective. In Proceedings of the 6th International Conference on Cartography and GIS, Albena, Bulgaria, 13–17 June 2016; pp. 482–492.
17. Touya, G.; Antoniou, V.; Olteanu-Raimond, A.M.; Van Damme, M.D. Assessing crowdsourced POI quality: Combining methods based on reference data, history, and spatial relations. *ISPRS Int. J. Geo-Inf.* **2017**, *63*, 80. [\[CrossRef\]](#)
18. Touya, G.; Antoniou, V.; Christophe, S.; Skopeliti, A. Production of Topographic Maps with VGI: Quality Management and Automation. In *Mapping and the Citizen Sensor*; Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.M., Fonte, C.C., Antoniou, V., Eds.; Ubiquity Press: London, UK, 2017; pp. 61–91. [\[CrossRef\]](#)
19. Zhang, L.; Pfoser, D. Using OpenStreetMap point-of-interest data to model urban change—A feasibility study. *PLoS ONE* **2019**, *14*, e0212606. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Fonte, C.C.; Antoniou, V.; Bastin, L.; Estima, J.; Arsanjani, J.J.; Bayas, J.C.L.; See, L.; Vatseva, R. Assessing VGI Data Quality. In *Mapping and the Citizen Sensor*; Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.M., Fonte, C.C., Antoniou, V., Eds.; Ubiquity Press: London, UK, 2017; pp. 137–163. [\[CrossRef\]](#)
21. Goodchild, M.F.; Li, L. Assuring the quality of volunteered geographic information. *Spat. Stat.* **2012**, *1*, 110–120. [\[CrossRef\]](#)
22. Meek, S.; Jackson, M.J.; Leibovici, D.G. A flexible framework for assessing the quality of crowdsourced data. In Proceedings of the 17th AGILE International Conference on Geographic Information Science, Castellón, Spain, 3–6 June 2014; Available online: https://agile-online.org/conference_paper/cds/agile_2014/agile2014_112.pdf (accessed on 16 March 2020).
23. Bordogna, G.; Carrara, P.; Criscuolo, L.; Pepe, M.; Rampini, A. A user-driven selection of VGI based on minimum acceptable quality levels. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *II-3/W5*, 277–284. [\[CrossRef\]](#)
24. Antoniou, V.; Skopeliti, A. Measures and indicators of VGI quality: An overview. In Proceedings of the ISPRS Geospatial Week 2015, La Grande Motte, France, 28 September–2 October 2015; pp. 345–351. Available online: <http://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/II-3-W5/345/2015/isprsannals-II-3-W5-345-2015.pdf> (accessed on 17 March 2021).
25. Senaratne, H.; Mobasher, A.; Ali, A.L.; Capineri, C.; Haklay, M. A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.* **2016**, *31*, 1–29. [\[CrossRef\]](#)
26. VGIscience Priority Programme. Available online: <https://www.vgiscience.org/about.html> (accessed on 16 May 2021).
27. Keil, J.; Edler, D.; Kuchinke, L.; Dickmann, F. Effects of visual map complexity on the attentional processing of landmarks. *PLoS ONE* **2020**, *15*, e0229575. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Skopeliti, A.; Antoniou, V.; Bandrova, T. Visualization and Communication of VGI Quality. In *Mapping and the Citizen Sensor*; Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.M., Fonte, C.C., Antoniou, V., Eds.; Ubiquity Press: London, UK, 2017; pp. 197–222. [\[CrossRef\]](#)
29. Kinkeldey, C.; MacEachren, A.M.; Schiewe, J. How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualization user studies. *Cartogr. J.* **2014**, *51*, 372–386. [\[CrossRef\]](#)

30. Skopeliti, A.; Antoniou, V.; Stamou, L. Visualisation and Communication of VGI Quality. In Proceedings of the 14th National Cartographic Conference, Thessaloniki, Greece, 2–4 November 2016. (In Greek).
31. Girres, J.F.; Touya, G. Quality assessment of the French OpenStreetMap dataset. *Trans. GIS* **2010**, *14*, 435–459. [\[CrossRef\]](#)
32. Arsanjani, J.J.; Barron, C.; Bakillah, M.; Helbich, M. Assessing the quality of OpenStreetMap contributors together with their contributions. In Proceedings of the 16th AGILE International Conference on Geographic Information Science, Leuven, Belgium, 14–17 May 2013; pp. 14–17.
33. Ali, A.L.; Schmid, F. Data quality assurance for Volunteered Geographic Information. In *Geographic Information Science, Proceedings of the 8th International Conference-GIScience 2014, Vienna, Austria, 24–26 September 2014*; Duckham, M., Pebesma, E., Stewart, K., Frank, A.U., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 126–141. ISBN 978-3-319-11593-1.
34. Sehra, S.S.; Singh, J.; Rai, H.S. A systematic study of OpenStreetMap data quality assessment. In Proceedings of the 11th International Conference on Information Technology: New Generations, Las Vegas, NV, USA, 7–9 April 2014; pp. 377–381. [\[CrossRef\]](#)
35. Touya, G.; Brando, C. Detecting Level-of-Detail inconsistencies in volunteered geographic information data sets. *Cartographica* **2013**, *48*, 134–143. [\[CrossRef\]](#)
36. Ávila Callau, A.; Pérez-Albert, Y.; Serrano Giné, D. Quality of GNSS Traces from VGI: A Data Cleaning Method Based on Activity Type and User Experience. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 727. [\[CrossRef\]](#)
37. Antoniou, V.; Skopeliti, A. The Impact of the Contribution Micro-environment on Data Quality: The Case of OSM. In *Mapping and the Citizen Sensor*; Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.M., Fonte, C.C., Antoniou, V., Eds.; Ubiquity Press: London, UK, 2017; pp. 165–196. [\[CrossRef\]](#)
38. Duchêne, C.; Ruas, A.; Cambier, C. The CartACom model: Transforming cartographic features into communicating agents for cartographic generalization. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1533–1562. [\[CrossRef\]](#)
39. Brando, C.; Bucher, B.; Abadie, N. Specifications for User Generated Spatial Content. In *Advancing Geoinformation Science for a Changing World, Springer-Verlag Lecture Notes in Geoinformation and Cartography*; Geertman, S., Reinhardt, W., Toppen, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2011.
40. Brando, C. Un Modèle d’ Operations Réconciliables Pour l’ Acquisition Distribuée de Données Géographiques. Ph.D. Thesis, Université Paris-Est, Champs-sur-Marne, France, 2013.
41. Servigne, S.; Ubeda, T.; Puricelli, A.; Laurini, R. A methodology for spatial consistency improvement of geographic databases. *Geoinformatica* **2000**, *4*, 7–34. [\[CrossRef\]](#)
42. Kainz, W. Spatial relationships-topology versus order. In Proceedings of the Fourth International Symposium on Spatial Data Handling, Zurich, Switzerland, 23–27 July 1990; pp. 814–819.
43. Martínez, P.; Martí, P.; Querin, O.M. Growth method for size, topology, and geometry optimization of truss structures. *Struct. Multidisc. Optim.* **2006**, *33*, 13–26. [\[CrossRef\]](#)
44. Papadias, D.; Mamoulis, N.; Delis, B. Algorithms for querying by spatial structure. In Proceedings of the 24th Very Large Data Bases Conference, New York, NY, USA, 24–27 August 1998; pp. 546–557.
45. Rodríguez, M.A.; Brisaboa, N.; Meza, J.; Luaces, M.R. Measuring consistency with respect to topological dependency constraints. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, New York, NY, USA, 2–5 November 2010; pp. 182–191. [\[CrossRef\]](#)
46. Brisaboa, N.; Luaces, M.; Rodríguez, M.A. Cognitive adequacy of topological consistency measures. In *International Conference on Conceptual Modeling*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 241–250.
47. Brisaboa, N.R.; Luaces, M.R.; Rodríguez, M.A.; Seco, D. An inconsistency measure of spatial data sets with respect to topological constraints. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 56–82. [\[CrossRef\]](#)
48. Open Street Map Wiki. Available online: <https://en.wikipedia.org/wiki/OpenStreetMap> (accessed on 18 March 2021).
49. Open Street Map. Available online: <https://www.openstreetmap.org/> (accessed on 18 March 2021).
50. Geofabrik. Available online: www.geofabrik.de (accessed on 18 March 2021).
51. BBBike. Available online: <https://download.bbbike.org/osm/bbbike/> (accessed on 18 March 2021).
52. Ramm, F. OpenStreetMap Data in Layered GIS Format Free Shapefiles. 2019. Available online: <https://download.geofabrik.de/osm-data-in-gis-formats-free.pdf> (accessed on 8 May 2021).
53. Nakos, B. The SP-Displacement Measure for Assessing Line Simplification. *Spat. Sci.* **2004**, *49*, 1–11. [\[CrossRef\]](#)
54. Davidovic, N.; Mooney, P.; Stoimenov, L.; Minghini, M. Tagging in volunteered geographic information: An analysis of tagging practices for cities and urban regions in OpenStreetMap. *ISPRS Int. J. Geo-Inf.* **2016**, *512*, 232. [\[CrossRef\]](#)
55. Mooney, P.; Minghini, M.; Laakso, M.; Antoniou, V.; Olteanu-Raimond, A.-M.; Skopeliti, A. Towards a Protocol for the Collection of VGI Vector Data. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 217. [\[CrossRef\]](#)
56. Zacharopoulou, D. Development of a GIS Tool for the Assessment of VGI Quality. Diploma Thesis, School of Rural and Surveying Engineering, National Technical University of Athens, Athens, Greece, October 2018.
57. Zacharopoulou, D. Evaluation and Visualization of Volunteered Geographic Information Consistency: Study of OpenStreetMap for European Cities. Master’s Thesis, School of Rural and Surveying Engineering, National Technical University of Athens, Athens, Greece, July 2020.
58. Urban Corridor Road Design: Guides, Objectives and Performance Indicators Sidewalks. Available online: https://www.roadspace.eu/wp-content/uploads/2019/11/MORE_D1_2_FINAL.pdf (accessed on 10 March 2021).

-
59. Roads. 2018. Available online: https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/ersosynthesis2018-roads.pdf (accessed on 10 March 2021).
 60. OSM Zoom Levels. Available online: https://wiki.openstreetmap.org/wiki/Zoom_levels (accessed on 13 May 2021).