

## OPENSTREETMAP AS AN INPUT SOURCE FOR PRODUCING GOVERNMENTAL DATASETS: THE CASE OF THE ITALIAN MILITARY GEOGRAPHIC INSTITUTE

A. Sarretta <sup>1</sup>, M. Napolitano <sup>2</sup>, M. Minghini <sup>3\*</sup>

<sup>1</sup> Research Institute for Geo-Hydrological Protection, National Research Council (CNR-IRPI), Padova, Italy – alessandro.sarretta@irpi.cnr.it

<sup>2</sup> Fondazione Bruno Kessler, Trento, Italy – napo@fbk.eu

<sup>3</sup> European Commission, Joint Research Centre (JRC), Ispra, Italy – marco.minghini@ec.europa.eu

**KEY WORDS:** Data integration, Data quality, Geospatial, Governmental dataset, Italian Military Geographic Institute, OpenStreetMap, Open data, Volunteered Geographic Information.

### ABSTRACT:

The role of Volunteered Geographic Information (VGI) to integrate, update or complement authoritative datasets released by governments has become increasingly important. This work analyses the contribution of OpenStreetMap (OSM), the most popular VGI project, as one of the input sources that the Military Geographic Institute (IGM), one of the Italian governmental mapping agencies, has used for producing the National Summary Database (DBSN). This database, which was recently released for 12 out of the 20 Italian regions, has a schema organised into a hierarchical structure composed of 10 layers, 30 themes and 93 classes, where each geospatial object carries information on the specific data source it was derived from. For each DBSN layer and theme, we first calculated the fraction of objects derived from OSM in all the Italian regions and related provinces. We found a heterogeneous picture with OSM contribution generally being limited, with the exception of few regions and layers/themes where the DBSN was almost exclusively derived from OSM. An in-depth comparison between the DBSN and OSM building datasets showed that OSM building completeness is varying across Italian regions and provinces, but in all regions there are buildings in OSM that are not included in the DBSN. The work shed light on the opportunities and obstacles for OSM to become a primary input source for the production of governmental datasets.

### 1. INTRODUCTION

The processes of collecting, curating, validating and publishing geospatial information have remained for centuries the sole prerogative of governmental public sector organisations. Geospatial data produced by governments have traditionally been considered as the reference source for the production of any other dataset or cartographic output. The potential of citizens as additional providers of geospatial data—termed Volunteered Geographic Information (VGI) in literature—materialised towards the beginning of our century and progressively challenged the traditional role of the public sector (Foody et al., 2017). In response to this shift, governments have soon started researching and exploring the use of VGI to update or complement their authoritative datasets—see e.g. Olteanu-Raimond et al. (2017), Minghini et al. (2019) and Jacquin et al. (2023)—and together with datasets from the private sector, satellites and research, VGI is currently considered a key element for the future generation of spatial data infrastructures (Kotsev et al., 2021).

Among the many types and sources of VGI, this paper focuses on OpenStreetMap (OSM, <https://openstreetmap.org>), initiated in 2004 and currently being the most successful VGI project. The goal of the project consists in the crowdsourced creation and update of a geospatial database of the whole world ([https://wiki.openstreetmap.org/wiki/Main\\_Page](https://wiki.openstreetmap.org/wiki/Main_Page)) that is made available under the open access Open Database License (ODbL). At the time of writing (May 2023), around 2 million people have contributed data to the OSM project (<https://wiki.openstreetmap.org/wiki/Stats>) and a full ecosystem of software tools, services and applications as well as a ded-

icated community have grown around it (Mooney and Minghini, 2017). In addition to the openly-licensed data, the success of OSM is mainly a consequence of its relatively simple data model, which makes it easy to both contribute data and consume them in standard Geographic Information Systems (GIS) applications. An overview of the OSM data model, which is out of scope for this paper, is provided by Minghini and Frassinelli (2019).

When investigating the use of OSM for integrating, complementing or updating authoritative datasets produced by governments, the main concern—being OSM a crowdsourced database—has been its quality. However, the huge amount of OSM quality assessments available in literature, both extrinsic (i.e. comparing OSM data with reference data such as governmental data, considered as ground truth) and intrinsic (i.e. only based on the analysis of OSM data, including the history of contributions) show that OSM is of comparable, and sometimes even better, quality than authoritative datasets—see e.g. Girres and Touya (2010); Haklay (2010); Fan et al. (2014); Antoniou and Skopeliti (2015); Fonte et al. (2016); Brovelli et al. (2016).

The actual integration between OSM and governmental datasets has also been explored in research, with authors conceptualising several conflation workflows for specific geospatial objects such as buildings, road networks, addresses or land use areas, and testing them at the city, regional (e.g. Fonte et al., 2017) or, more rarely, national level (e.g. Sarretta and Minghini, 2021). Outside the research context, the most successful examples of integration between OSM and governmental data are OSM imports, where external datasets available under licenses compatible with the OSM's ODbL are imported into OSM, typically by regional or national OSM communities. The list of completed, ongoing and planned OSM imports is maintained

\* Corresponding author

by the OSM community at <https://wiki.openstreetmap.org/wiki/Import/Catalogue>. In some few cases, mutually-beneficial partnerships are established between governmental organisations and OSM communities: the former release their datasets so that they can be imported in OSM; the latter further improve/update the OSM database (e.g. by reflecting changes occurring over time) and the former extract back the OSM data to improve or update their national datasets. This is for example the case of the governmental address database in the Netherlands (Granell et al., 2022). While the integration between OSM and authoritative datasets is typically feasible from a technical point of view, sometimes the real barrier that prevents it is the license of the governmental dataset, which may not be compatible with the ODbL. Examples of such situations happened at the French and Spanish national mapping agencies and are described by Granell et al. (2022).

Recent years have also witnessed an increased use of OSM by the private sector, including by some of the world's largest companies offering digital services such as Meta (formerly Facebook), Amazon, Apple and Microsoft. In exchange of leveraging OSM data to foster their business, such companies typically contribute back to OSM: economically (for example by sponsoring the annual OSM conferences), by hiring staff that improves the database, and—which is of major interest in this context—by publicly releasing datasets generated from OSM. This is for example the case of Microsoft's building footprints, a dataset of around 1 billion objects extracted from OSM data and Bing Maps imagery from 2014 to 2022 (<https://www.microsoft.com/en-us/maps/building-footprints>). In December 2022 Amazon, Meta, Microsoft and TomTom established the Overture Maps Foundation (<https://overturemaps.org>), which promised to release and curate quality-checked worldwide map data from the aggregation of multiple open data sources, including from governments, civic organisations and OSM.

In this complex context where OSM is increasingly used as a key input dataset by several organisations from the public and private sector, this paper analyses the case of the Italian Military Geographic Institute (*Istituto Geografico Militare*, IGM). IGM is one of the Italian governmental mapping agencies and is responsible for the production of maps at the scale 1:25,000. In September 2022 the IGM released a new product, namely the National Summary Database (*Database di Sintesi Nazionale*, DBSN), which explicitly mentions OSM among the input data sourced used for its production (<https://www.igmi.org/en/dbsn-database-di-sintesi-nazionale>). Based on the analysis of this authoritative database, the paper aims to identify the actual role played by OSM in its production process as well as the reasons for such choice.

The remainder of the paper is structured as follows. Section 2 describes the IGM's DBSN database in detail and briefly introduces OSM, identifying the objects of interest for the study. Section 3 presents the analysis on the DBSN, whose results are discussed in Section 4. The paper concludes with Section 5, which offers some reflections on the results and implications for future work.

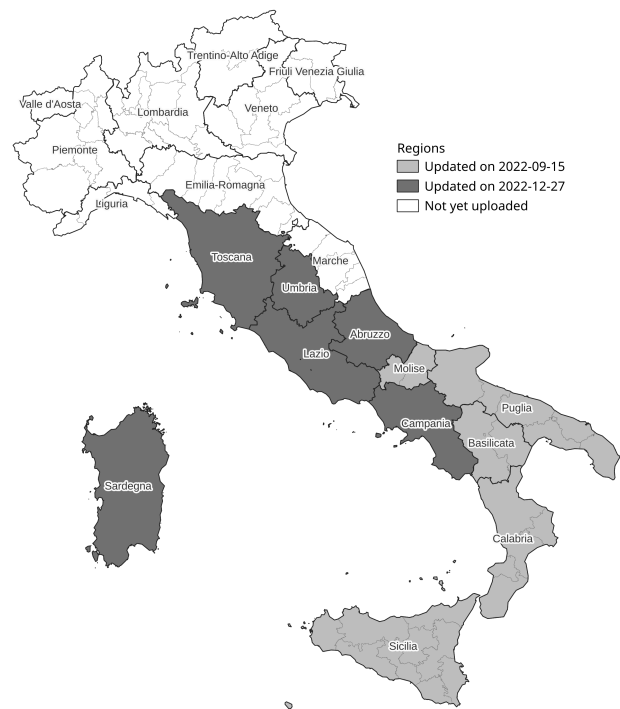
## 2. DATASETS

As introduced in Section 1, this work considers two geospatial datasets: the National Summary Database produced by the

Italian IGM, and OSM. While sharing a broad spectrum of thematic domains, they have a very different genesis process and management approach. The following sub-sections introduce the two datasets and describe their content and structure.

### 2.1 National Summary Database (DBSN)

The National Summary Database (*Database di Sintesi Nazionale*, DBSN) is a vector database of geospatial information relevant for analysis and representation at the national level, which is also used to derive maps at the scale 1:25,000 through automatic procedures. It has been released by the Italian Military Geographic Institute (IGM) starting from September 2022 and currently includes data covering 12 out of the 20 Italian regions (Abruzzo, Basilicata, Calabria, Campania, Lazio, Marche, Molise, Puglia, Sardegna, Sicilia, Toscana and Umbria). These regions are all located in central and southern Italy (see Figure 1). Data for the remaining 8 regions will be released in the near future. IGM is one of the Italian governmental mapping agencies and has played a central role in the production and management of official cartographic products in Italy since 1861.



**Figure 1.** Map showing the availability and date of last update of the IGM's DBSN for the 20 Italian regions.

The DBSN schema is a subset of the specifications defined in the document "Catalogue of Spatial Data — Content Specifications for Geotopographic Databases" (Agency for Digital Italy, 2015), annexed to the Ministerial Decree of 10 November 2011 "Technical rules for the definition of content specifications of geotopographical databases" (Presidency of the Council of Ministers, 2012). The DBSN schema has a hierarchical structure composed of 10 layers, 30 themes and 93 classes. Table 1 lists the layers and corresponding themes.

Upon registration, data are freely available for download from a dedicated page of the IGM website (<https://www.igmi.org/it/dbsn-database-di-sintesi-nazionale>). Datasets are available as geodatabase files zipped for each province (with 60 provinces currently available) and their date of last update

DBSN layer	DBSN theme
Geodetic and photogrammetric information	Geodetic information
	Cartographic and meta-information
Roads, mobility and transport	Roads
	Railways
	Other transport
Buildings and human settlements	Buildings
	Man-made constructions
	Transport infrastructure works
	Soil support and defence works
	Hydraulic, defence and hydraulic regulation works
Hydrography	Inland and transitional waters
	Marine waters
	Glaciers and perennial snow-fields
	Hydrographic network
Orography	Altimetry
	Bathymetry
	Terrain forms
	Digital terrain models (tin, dem/dtm)
Vegetation	Agro-forestry areas
	Urban green
Underground utility networks	Water supply network
	Electricity network
	Gas distribution network
	Oil pipelines
Significant places and cartographic markings	Significant places
	Cartographic markings
Administrative areas	Local authority administrative areas
Appurtenant areas	Transport services
	Appurtenances
	Quarries-dumps

**Table 1.** List of layers and themes as defined in the DBSN.

is: 15 September 2022 for provinces in the regions Basilicata, Calabria, Molise, Puglia and Sicilia; and 27 December 2022 for provinces in the regions Abruzzo, Campania, Lazio, Marche, Sardegna, Toscana and Umbria (see Figure 1).

The creation of the DBSN leverages a number of information sources, with regional geotopographic databases being the primary source of information, and products from other national public sector bodies (such as cadastral maps) as additional ones. The source of each object in the DBSN database is recorded in a specific attribute field (*meta\_list*) through a list of codes corresponding to the various sources (see Table 2). Among the external sources used as input for the production of the DBSN, OSM is explicitly considered.

Source code	Source institution	Explanation
01	<i>igm</i>	IGM
02	<i>ortofoto AGEA</i>	AGEA orthophoto
03	<i>osm</i>	OpenStreetMap
04	<i>db_regionale</i>	regional database
...	...	...

**Table 2.** First 4 codes, names of the data sources (as stored in the *meta\_list* field of the DBSN database) and corresponding explanation.

An element of novelty, at least in the Italian context, is the release of the DBSN under the ODbL license. This is a consequence of the fact that the ODbL, under which OSM is li-

DBSN theme	OSM tags
Roads	highway=*
Railways	railway=*
Other transport	aerialway=*, route=ferry, ...
Buildings	building=*
Man-made constructions	man_made=*
Transport infrastructure works	bridge=*, highway=* + tunnel=*, ...
Soil support and defence works	man_made=embankment
Hydraulic, defence and hydraulic regulation works	waterway=dam, man_made=dyke, embankment=dyke

**Table 3.** DBSN themes belonging to the layers *Roads, mobility and transport* and *Buildings and human settlements* and corresponding (not exhaustive) OSM tags.

censed, requires derivative products to be released under the same license (share-alike clause). The DBSN database was downloaded from the IGM website in April 2023.

## 2.2 OpenStreetMap (OSM)

The data structure in OSM is based on three main geometric primitives, namely nodes, ways and relations (<https://wiki.openstreetmap.org/wiki/Elements>) and a set of tags characterising the semantic component. Tags describe the attributes of OSM objects using an unlimited list of key-value pairs in a simple and modular structure that is easy to update and enrich (Minghini and Frassinelli, 2019).

A preliminary matching table between the DBSN data structure and the OSM tagging scheme was provided by the IGM representatives to the authors during some meetings organised in preparation of this research work. A complete analysis of the correspondence between the two databases is out of the scope of this work. However, by way of example Table 3 shows DBSN themes and their corresponding OSM tags for two DBSN layers: *Roads, mobility and transport* and *Buildings and human settlements*. OSM data was downloaded on 20 April 2023 using the procedure described in Section 3.

## 3. METHODOLOGY

The DBSN and OSM databases differ in many respects, including the overall purpose, data specifications, governance, data collection and management process, spatial scale and update frequency. Despite these differences, both databases have the common characteristic of not focusing on one specific theme, since they both contain objects belonging to several domains. For this reason, in order to perform a comprehensive analysis at the national level, we started with a general investigation of the OSM contribution as a source of information for the DBSN for all its themes and classes. Afterwards, we compared in more depth the DBSN and OSM databases with an exclusive focus on buildings (*Buildings* theme). These are among the most relevant datasets in several map-based applications; they have been recently included among the so-called “high-value datasets” that European Union Member States public sector organisations shall make available for free and under open licenses (European Commission, 2023).

The first analysis was performed using only the DBSN, specifically the attribute *meta\_list* (see Sub-section 2.1). Instead,

the second analysis required the combined use of the DBSN and OSM databases. The analyses were performed through a set of Python scripts, which are licensed under the open source WTFPL and made available at <https://github.com/napo/dbsnosmcompare>. The analyses consist of six steps, each of which is associated with a Python script. These scripts are also available in the Jupyter Notebook format to enable custom interactive reworking. In addition to ensuring reproducibility of results, the purpose is also to produce tools for replicating the analyses on future releases of the DBSN database, e.g. for the Italian regions that are not yet included. The primary outputs of the analyses are .csv and .parquet files that include the results (see Section 4) and graphs.

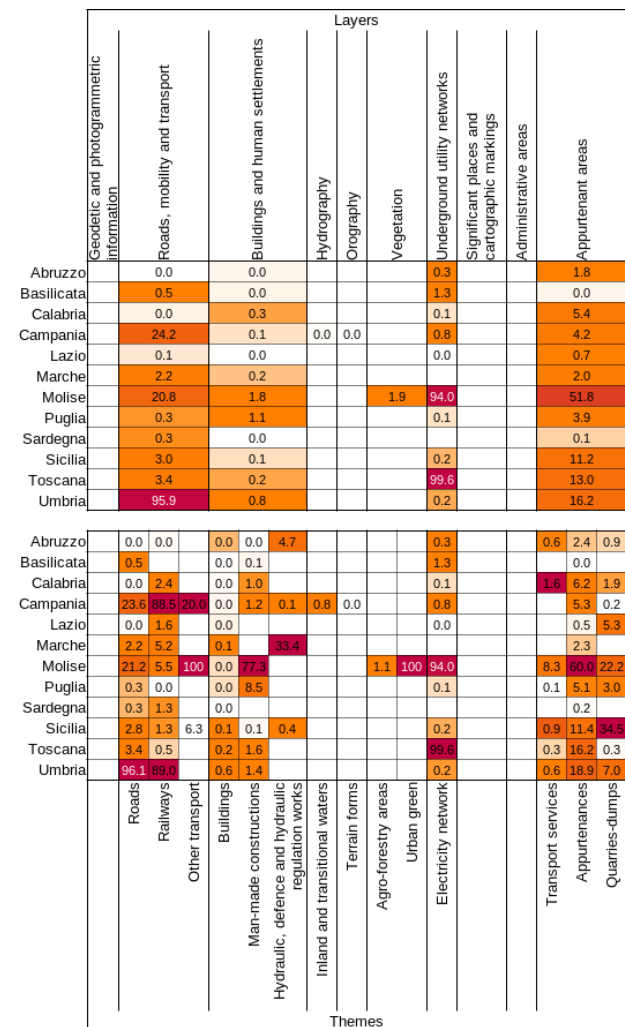
The first step involves downloading the DBSN and OSM datasets. Regarding the former, the IGM makes use of a web application to manage data distribution by tracking users and payments. Although the DBSN is an open dataset available for free, the IGM still chose to use this platform, without however requiring end-users to pay for the dataset. The only constraint is that users have to register in order to download the dataset. Nevertheless, once downloaded the dataset can be re-distributed under the (permissive) conditions required by the ODbL. As a result, the Italian OSM community created a wiki page with direct links to download the DBSN files (<https://wiki.openstreetmap.org/wiki/Italy/DBSN>). IGM data are made available in the geodatabase format (.gdb). OSM data were downloaded from the “Italian OpenStreetMap Extracts” (<https://osmit-estratti.wmcloud.org>), a web application provided by Wikimedia Italia, the Italian local chapter of the OpenStreetMap Foundation. This web application publishes, on a daily basis, OSM data for each Italian municipality, province and region (the three main administrative levels). OSM data are provided in both the Protocolbuffer Binary Format (.pbf) and the GeoPackage (.gpkg) format. The latter, which essentially derives from a conversion of the .pbf file into an encoding suitable for consumption in GIS tools, converts OSM primitives into points, lines, and polygons with the respective tags.

The second step involves identifying, for each region and for each province, the features of the DBSN database derived from OSM, by filtering out those with the *osm* value of the *meta\_list* field (see Table 2). In the third step, the filtered data is further enriched by including the additional attributes provided by the IGM, which are converted in tabular form and translated into English. In the fourth step, results are aggregated to measure the degree to which the IGM makes use of OSM data to produce the DBSN, for each province and for each region. The output of this process consists of heatmap tables and .csv files that can be reused for further analysis. The last two steps perform a comparison between the DBSN and OSM datasets, focused on buildings. The OSM database is first clipped on the boundary of each province and region (retrieved from the DBSN). Then, the total areas of buildings are computed for both datasets for each province and region. In addition to a comparison between the DBSN and OSM datasets, this may also provide an indication on the completeness of OSM buildings. Finally, a spatial intersection between the two datasets makes it possible to compute the fractions of the areas of OSM buildings that overlap and do not overlap with the buildings in the DBSN.

#### 4. RESULTS

The contribution of OSM as a source of information for the DBSN is highly variable among the 12 available regions. Fig-

ure 2 shows, for all regions, the percentage of DBSN features derived from OSM for each DBSN layer (top) and theme (bottom). Values equal to 0% (corresponding to the absence of OSM contribution) appear as blank cells, while values shown as “0.0” correspond to percentages lower than 0.1% but still higher than 0%. Only 7 out of the 10 DBSN layers include some contributions from OSM (see the top half of Figure 2). Four of them (*Roads, mobility and transport; Buildings and human settlements; Underground utility networks* and *Appurtenant areas*) show a relative wide presence of OSM across all regions; in contrast, OSM is only used in one region for *Vegetation* (Molise) and for *Hydrography* and *Orography* (Campania). In a few cases (namely *Roads, mobility and transport* in Umbria, and *Underground utility networks* in Molise and Toscana), OSM is the source of information for more than 90% of the DBSN features. However, in almost half of the cases, the contribution of OSM is extremely limited, with values lower than 1%. The layers where OSM does not contribute at all to the DBSN are *Geodetic and photogrammetric information, Significant places and cartographic markings* and *Administrative areas*.



**Figure 2.** Percentage of DBSN objects derived from OSM for the available regions, for each and DBSN layer (top half) and theme (bottom half). Colours correspond to percentage classes, from white for low values to dark red for high values.

The bottom half of Figure 2 shows the percentage of DBSN objects derived from OSM, for the themes containing at least

one OSM-derived object. This means that 16 out of the total 30 themes defined in the DBSN schema (see Table 1) do not contain any feature derived from OSM and therefore are not represented in Figure 2. These are *Geodetic information; Cartographic and meta-information; Transport infrastructure works; Soil support and defence works; Marine waters; Glaciers and perennial snowfields; Hydrographic network; Altimetry; Bathymetry; Digital terrain models (tin, dem/dtm); Water supply network; Gas distribution network; Oil pipelines; Significant places; Cartographic markings and Local authority administrative areas*. The percentage values for DBSN themes confirm the heterogeneous picture identified for layers, with few regions (mainly Campania, Molise and Umbria) being significantly dependent on OSM, at least for some themes, and most regions only showing little contributions from OSM. However, it should be noted that the percentage values shown in Figure 2 strongly depend on the total number of objects available for each DBSN layer and theme, i.e. a high percentage does not necessarily correspond to several OSM objects being used. As an example, in the Molise region the theme *Other transport* is fully derived from OSM (100% value), but this actually corresponds to only one object (a chairlift in the province of Isernia). Figure 3 further zooms at the scale of Italian provinces and shows for each of them the percentage of DBSN objects derived from OSM, for the themes containing at least one OSM-derived object. In most cases, all provinces within the same region show a similar degree of dependence on OSM, thus fully reflecting the regional trend depicted in Figure 2. However, in some cases only one or few provinces within the same region show, for the same DBSN theme, the integration of OSM objects. For example, for the *Other transport* theme the dependence on OSM exclusively happens in the provinces of Napoli for Campania region and Isernia for Molise region.

As described in Section 3, a more in-depth comparison between the DBSN and the OSM datasets was then performed on buildings. Figure 4 classifies the Italian provinces (belonging to the 12 regions where the DBSN is available) according to the ratio, expressed as a percentage, between the total area of buildings in OSM (extracted as shown in Table 3) and in the DBSN. The overall picture is once again heterogeneous. Only in three regions (Puglia, Toscana and Sardegna) the total area of OSM buildings is almost equal (and in some cases, even higher) than the total area of buildings in the DBSN for all provinces. In all the other regions the classification varies among provinces, including cases where within the same region there are both provinces with very high and very low percentages (e.g. Lazio). This analysis provides also an indirect measure of the completeness of OSM buildings across Italian provinces. This shows significant local variations that may be due to different reasons, including: (i) the demographic density (more densely populated areas have a higher chance to be mapped); (ii) the attractiveness (more touristic areas attract more people and have a higher chance to be mapped); and (iii) the presence or absence of OSM local communities that add and update OSM objects.

Table 4 complements Figure 4 with a summary on the comparison between the OSM and DBSN building datasets, this time performed at the level of Italian regions. The second column shows the percentage ratio between the total areas of buildings in the OSM and DBSN datasets. As expected, the highest values occur for Puglia, Toscana and Sardegna, while the minimum value (for Calabria region) is 35.4%. The average value for the 12 regions (not included in Table 4) is 63.7%. This confirms the variability of completeness of the OSM building data-

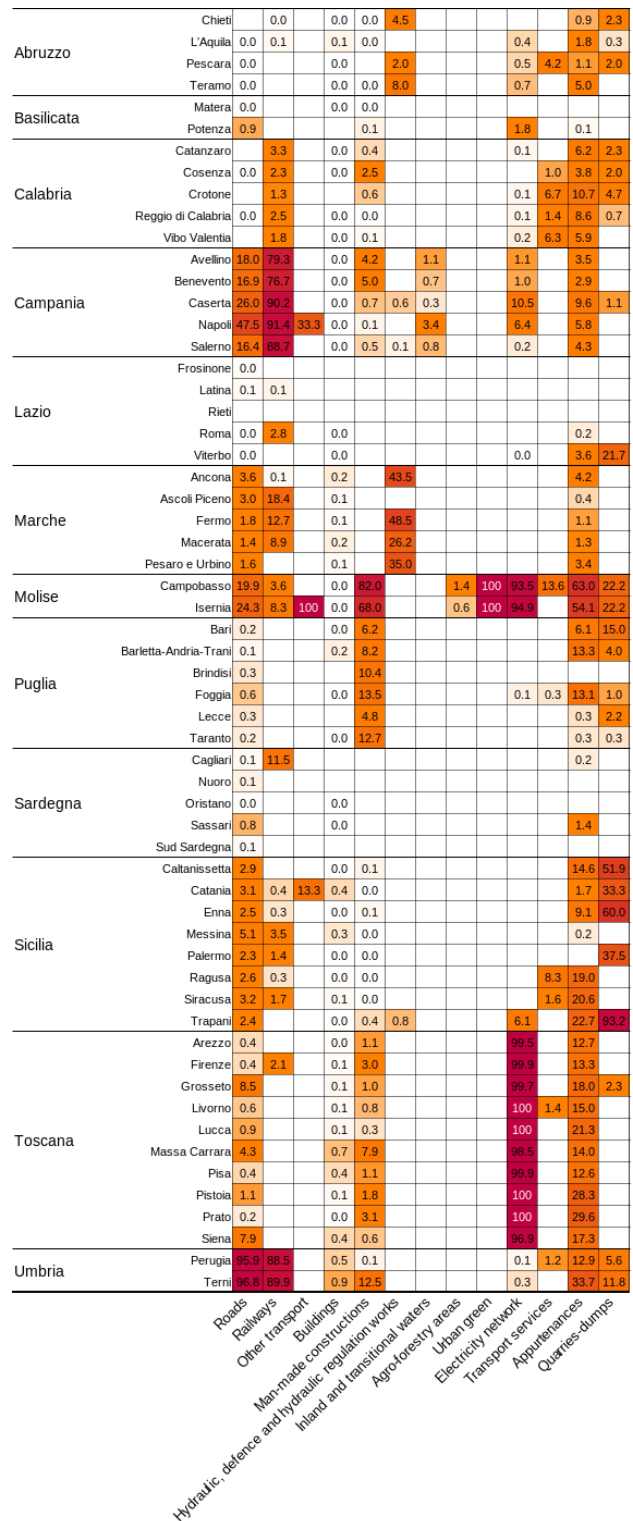
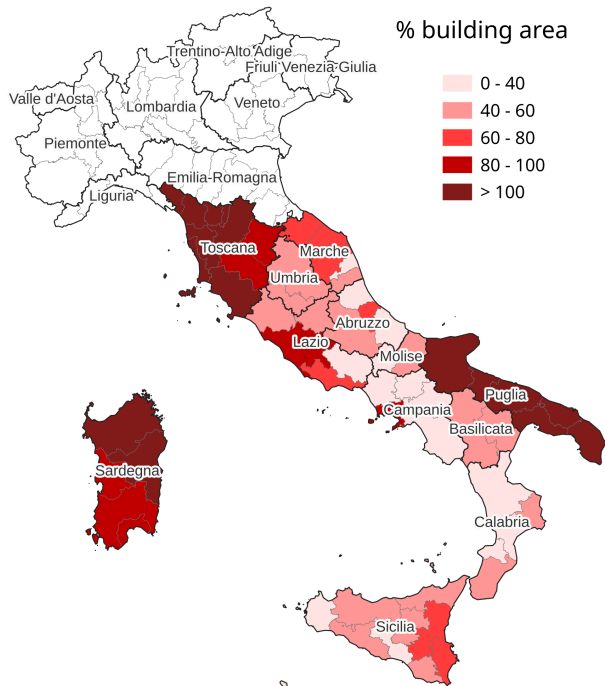


Figure 3. Percentage of DBSN objects derived from OSM for the available provinces, for each DBSN theme containing at least one OSM-derived object. Colours correspond to percentage classes, from white for low values to dark red for high values.

sets, even at the regional level. The third column of Table 4 shows the standard deviation of the percentage ratio between the total areas of the OSM and DBSN buildings, computed for the provinces of each region. As already suggested from Figure 4, the variability between provinces in the same region shows



**Figure 4.** Classification of Italian provinces based on the ratio between the total area of buildings in OSM and in the DBSN.

values from very low (the minimum is 1.7% for Puglia, where percentage ratios range from 102.3% to 107.4%) to very high (the maximum is 30.8% for Campania, where percentage ratios range from 18.8% for Caserta to 99.7% for Napoli). Finally, for each region we performed a spatial intersection between the OSM and DBSN building datasets and computed the percentage of the area of OSM buildings that do not intersect any building in the DBSN dataset. These values, shown in the last column of Table 4, are all lower than 4% with the exceptions of Marche (6.7%), Puglia (6.4%) and Sardegna (5.2%). This however proves that OSM includes some buildings, which are not present in the governmental dataset produced by the IGM.

## 5. DISCUSSION AND CONCLUSIONS

The ever-increasing need to produce new geospatial datasets and to update existing ones has led one of Italy's leading governmental map producers, the IGM, to rethink the way cartographic outputs should be created, moving away from the traditional approach of producing them from scratch in favour of leveraging the aggregation of already available authoritative (e.g. regional and cadastral) and non-authoritative data sources, the latter including OSM. This paper offered a first analysis of the IGM's recently released dataset, the DBSN. Currently available for 12 out of the 20 Italian regions, it makes use of the authoritative databases of these regions as the primary input source, whose deficiencies and gaps were then complemented (with a variable degree) by additional sources. In particular, OSM is mainly used for DBSN data categories related to transportation, buildings and utility networks, which represent baseline cartographic information. In contrast geodetic, photogrammetric, hydrographical, orographical and administrative information is almost never derived from OSM. For most of the categories of information where OSM is used, it usually plays a minor role as an input source. However, in few regions most

Region	% OSM/DBSN building area	St. dev. of % OSM/DBSN building area for provinces	% OSM building area disjoint from DBSN
Abruzzo	44.2	9.9	2.9
Basilicata	46.8	7.4	3.1
Calabria	35.4	11.6	2.2
Campania	50.6	30.8	3.1
Lazio	71.9	22.2	3.9
Marche	57.8	16.7	6.7
Molise	47.6	7.8	3.1
Puglia	105.2	1.7	6.4
Sardegna	99.1	4.1	5.2
Sicilia	50.0	11.9	2.3
Toscana	103.7	5.2	7.1
Umbria	51.9	7.2	2.0

**Table 4.** Percentage ratio between the total areas of OSM and DBSN buildings for each region; standard deviation of the percentage ratio between the total areas of the OSM and DBSN buildings, for the provinces of each region; percentage of the area of OSM buildings disjoint from DBSN buildings for each region.

(and in some cases, all) of the DBSN objects for specific categories are derived from OSM, e.g. for the transportation network in Umbria and the underground utility networks in Molise and Toscana. Albeit limited to buildings only, the results also suggest that some objects available in OSM are (still) not included in the DBSN. There might be at least two reasons for this: (i) the current workflow adopted by the IGM to filter the OSM database (through tags) does not capture potentially relevant objects; and (ii) the continuous evolution of the OSM database happens at a pace that the IGM can hardly cope with. The latter is a general consideration applying to any governmental organisation willing to include OSM in its data gathering workflow, which, in an ideal case, would require the setup of automated processing pipelines to regularly fetch OSM updates and include them in the final product.

This study confirms the importance achieved by OSM and the role it currently plays as a reference source of geospatial information for governmental bodies. Compared to other cases where the actual integration of OSM into authoritative datasets was not possible because of their license incompatibility (Saretta and Minghini, 2021; Granell et al., 2022), the IGM case represents a significant example of how a government selects the license for its dataset to explicitly allow OSM integration.

Stemming from the results of this work, we can make some conclusions on the OSM contribution for integration into Italian authoritative datasets that are likely to be valid in general. First, the comparison between the DBSN and OSM building datasets shows that the completeness of the latter is heterogeneous across the country. This is well-known from literature and means that OSM, while already representing a useful asset to update or complement authoritative datasets, would still need some effort to become a primary source for their production. This calls for an intervention from OSM communities at the national and regional levels to improve, in a coordinated manner (including through imports), categories of OSM objects that are most in need, often within a well-defined geographical area. In Italy, successful initiatives of this kind have already taken place in the past, e.g. to address the COVID-19 emergency (Minghini et al., 2022). Failures from OSM 'alone' to fill the existing data gap would favour initiatives from the private sector (e.g. the one from the Overture Maps Foundation), which can leverage huge economic and staff resources to create new, complete and

quality-checked datasets derived from OSM, including at the national and even global level.

Future work can be developed along multiple directions. First, the same analyses could be replicated for the remaining 8 Italian regions once the DBSN will be released for them. In addition to providing the full national picture on the comparison between the DBSN and the OSM datasets and possibly identifying some geographical trends, results may reveal other OSM objects having a high(er) potential for integration in the DBSN. In addition, analyses similar to the one presented for buildings could be performed for other DBSN layers and themes. At the time of writing, an investigation on the *Roads* DBSN theme is already ongoing. Finally, the degree to which OSM is used for the production of the DBSN may depend on both the quality of OSM itself (of which completeness, addressed in this paper, only represents one element) and the quality of the regional authoritative databases. Hence, future work could focus on analysing: (i) the correlation between the OSM quality and the OSM use within the DBSN across provinces and regions; and (ii) the comparison between the OSM and the authoritative databases for each Italian region, and how the results of this relate to the use of OSM for the DBSN production.

#### DISCLAIMER

The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.

#### REFERENCES

- Agency for Digital Italy, 2015. Catalogo dei dati territoriali — Specifiche di contenuto per i DataBase Geotopografici. [https://geodati.gov.it/geoportale/images/Specifica\\_GdL2\\_09-05-2016.pdf](https://geodati.gov.it/geoportale/images/Specifica_GdL2_09-05-2016.pdf) (25 May 2023).
- Antoniou, V., Skopeliti, A., 2015. Measures and indicators of VGI quality: an overview. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, II-3/W5. <https://doi.org/10.5194/isprsannals-II-3-W5-345-2015>.
- Brovelli, M. A., Minghini, M., Molinari, M. E., Zamboni, G., 2016. Positional accuracy assessment of the OpenStreetMap buildings layer through automatic homologous pairs detection: The method and a case study. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B2, 615–620. <https://doi.org/10.5194/isprsarchives-XLI-B2-615-2016>.
- European Commission, 2023. Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use. *Official Journal of the European Union*, L 19, 43–75.
- Fan, H., Zipf, A., Fu, Q., Neis, P., 2014. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28(4), 700–719. <https://doi.org/10.1080/13658816.2013.867495>.
- Fonte, C. C., Minghini, M., Patriarca, J., Antoniou, V., See, L., Skopeliti, A., 2017. Generating up-to-date and detailed land use and land cover maps using OpenStreetMap and GlobeLand30. *ISPRS International Journal of Geo-Information*, 6(4), 125. <https://doi.org/10.3390/ijgi6040125>.
- Fonte, C., Minghini, M., Antoniou, V., See, L., Patriarca, J., Brovelli, M., Milcinski, G., 2016. An automated methodology for converting OSM data into a land use/cover map. *Proceedings of the 6th International Conference on Cartography and GIS*.
- Foody, G., Fritz, S., Fonte, C. C., Bastin, L., Olteanu-Raimond, A.-M., Mooney, P., See, L., Antoniou, V., Liu, H.-Y., Minghini, M., Vatsava, R., 2017. Mapping and the Citizen Sensor. G. Foody, L. See, S. Fritz, P. Mooney, A.-M. Olteanu-Raimond, C. C. Fonte, V. Antoniou (eds), *Mapping and the Citizen Sensor*, Ubiquity Press, 1–12. <https://doi.org/10.5334/bbf.a>.
- Girres, J.-F., Touya, G., 2010. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4), 435–459. <https://doi.org/10.1111/j.1467-9671.2010.01203.x>.
- Granell, C., Mooney, P., Jirka, S., Rieke, M., Ostermann, F., van Den Broecke, J., Sarretta, A., Verhulst, S., Dencik, L., Oost, H., Micheli, M., Minghini, M., Kotsev, A., Schade, S., 2022. Emerging approaches for data-driven innovation in Europe: Sandbox experiments on the governance of data and technology.
- Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning and design*, 37(4), 682–703. <https://doi.org/10.1068/b35097>.
- Jacquin, C., Gómez, P. M., Arriaga, V., Santiago, A., 2023. Citizen science interactions with official geospatial information; Case studies from Mexico. *Frontiers in Environmental Science*, 10. <https://doi.org/10.3389/fenvs.2022.800247>.
- Kotsev, A., Minghini, M., Cetl, V., Penninga, F., Robbrecht, J., Lutz, M., 2021. INSPIRE: A Public Sector Contribution to the European Green Deal Data Space. EUR 30832 EN, Publications Office of the European Union, Luxembourg. <https://doi.org/10.2760/8563>.
- Minghini, M., Frassinelli, F., 2019. OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date? *Open Geospatial Data, Software and Standards*, 4(1), 1–17. <https://doi.org/10.1186/s40965-019-0067-x>.
- Minghini, M., Kotsev, A., Lutz, M., 2019. Comparing INSPIRE and OpenStreetMap data: how to make the most out of the two worlds. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W, 167–174. <https://doi.org/10.5194/isprs-archives-XLII-4-W14-167-2019>.
- Minghini, M., Sarretta, A., Napolitano, M., 2022. OpenStreetMap contribution to local data ecosystems in COVID-19 times: Experiences and reflections from the Italian case. *Data*, 7(4), 39. <https://doi.org/10.3390/data7040039>.
- Mooney, P., Minghini, M., 2017. A review of OpenStreetMap data. G. Foody, L. See, S. Fritz, P. Mooney, A.-M. Olteanu-Raimond, C. C. Fonte, V. Antoniou (eds), *Mapping and the*

*Citizen Sensor*, Ubiquity Press, 37–59. <https://doi.org/10.5334/bbf.c>.

Olteanu-Raimond, A.-M., Laakso, M., Antoniou, V., Fonte, C. C., Fonseca, A., Grus, M., Harding, J., Kellenberger, T., Minghini, M., Skopeliti, A., 2017. VGI in National Mapping Agencies: Experiences and Recommendations. G. Foody, L. See, S. Fritz, P. Mooney, A.-M. Olteanu-Raimond, C. C. Fonte, V. Antoniou (eds), *Mapping and the Citizen Sensor*, Ubiquity Press, 299–326. <https://doi.org/10.5334/bbf.m>.

Presidency of the Council of Ministers, 2012. Regole tecniche per la definizione delle specifiche di contenuto dei database geotopografici. Ministerial Decree 10 November 2011, Gazzetta Ufficiale 48 of 27/02/2012, <https://www.gazzettaufficiale.it/eli/id/2012/02/27/12A01800> (25 May 2023).

Sarretta, A., Minghini, M., 2021. Towards the integration of authoritative and OpenStreetMap geospatial datasets in support of the European strategy for data. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-4-W2-2021, 159–166. <https://doi.org/10.5194/isprs-archives-XLVI-4-W2-2021-159-2021>.