European Commission

JRC TECHNICAL REPORT

EU gazetteer evaluation

CETL, Vlado
VREČAR, Simon
REUTER, Hannes Isaak
BOGUSLAWSKI, Ray
PIGNATELLI, Francesco

2020

ELISE

Enabling digital government through geospatial & location intelligence

Joint Research Centre

EUR 30378 EN

Contact information

Name: Francesco Pignatelli
Address: European Commission, Joint Research Centre, TP 263, Via E. Fermi 2749, 21027 Ispra (VA), Italy
Email: francesco.pignatelli@ec.europa.eu
Tel.: +39 0332 78 6319

Name: Vlado Cetl
Address: European Commission, Joint Research Centre, TP 263, Via E. Fermi 2749, 21027 Ispra (VA), Italy
Email: vlado.cetl@ec.europa.eu
Tel.: +39 0332 78 3026

EU Science Hub

https://ec.europa.eu/jrc

JRC121541

EUR 30378 EN

PDF ISBN 978-92-76-22301-6 ISSN 1831-9424 doi:10.2760/095025

Luxembourg: Publications Office of the European Union, 2020

How to cite this report: Cetl, V., Vrečar, S., Reuter, H. I., Boguslawski, R., Pignatelli, F., *EU gazetteer evaluation*, EUR 30378 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-22301-6, doi:10.2760/095025, JRC121541.

Contents

Acknowledgements

We thank our colleague Marco Minghini for his thoughtful constructive feedback that greatly improved the report.

Authors

CETL, Vlado	European Commission, DG Joint Research Centre
VREČAR, Simon	European Commission, DG Joint Research Centre, External Consultant
REUTER, Hannes Isaak	European Commission, DG Eurostat
BOGUSLAWSKI, Ray	European Commission, DG Joint Research Centre, External Consultant
PIGNATELLI, Francesco	European Commission, DG Joint Research Centre

Abstract

This JRC technical report summarises the ELISE (European Location Interoperability Solutions for e-Government) activities in support to the development of an EU gazetteer. Most Member States have their own national gazetteer service so, if an EU gazetteer service is to be justified, there needs to be sufficient demand for pan-European applications or sufficient added value beyond existing national gazetteers. The ELISE Action of the ISA^2 Programme carried out a survey in conjunction with EuroGeographics in 2018, aimed at understanding the demand-side and supply-side perspectives related to pan-European gazetteer data and services. The results clearly showed that there is demand for an EU gazetteer to support multi-national applications or complement existing national gazetteers, for purposes such as emergency response, searching for datasets, news items, or tourism / cultural heritage sites, validating foreign addresses, etc. This report further investigates two datasets on the pan-European level: Geographical names and Addresses as the most relevant datasets for the EU gazetteer. In the report we also analyse authoritative vs. volunteered spatial datasets. The results of the analysis showed that both data sources, official and volunteered, are complementary and mutually enhanced results can be obtained by combining the two. In addition, "Cultural Heritage Testbed" application has been developed with the aim to identify data, functionality gaps and improvements needed in different gazetteer solutions. The findings and possible applications were discussed with several existing use cases, with cross-border and pan-European coverage. Overall findings in this report can be used to justify the relevance and importance of Geographical names and Addresses datasets in the context of defining future high value datasets at an EU level.

1 Introduction

1.1 European Location Interoperability Solutions for e-Government (ELISE)

The Digital Economy Unit of the EC Joint Research Centre (JRC), in cooperation with other services of the European Commission, is coordinating the "European Location Interoperability Solutions for e-Government (ELISE[1])", Action 4.1 of the ISA[2] Programme.

The ELISE Action is a package of legal/policy, organisational, semantic and technical interoperability solutions to facilitate efficient and effective electronic cross-border or cross-sector interaction between European public administrations and between them and citizens and businesses, in the domain of location information and services, supporting Digital Single Market (DSM), Better Regulation (BR) and Public Sector Modernisation (PSM) goals. It is aligned with the focus of ISA^2 on European public administrations, businesses and citizens, and the need to ensure that best practice interoperable solutions are deployed across the European Union (EU).

ELISE continues and builds on the work of the European Union Location Framework (EULF[3]) and A Reusable INSPIRE Reference Platform (ARE3NA[4]) Actions in the ISA[5] programme, which partially addressed the challenges and opportunities in location-related interoperability, in terms of frameworks, application pilots and re-usable tools.

1.2 Context

ISA^2 and previously ISA geospatial stakeholders have recognised the need for ready access to reference datasets that are central to many applications. These include geographical names, administrative units, addresses, and buildings datasets that are typically used in "gazetteer" services that establish the link between geographical names and their locations, as well as many other applications.

A gazetteer is a register of features of a country, region, continent, etc. containing information on their geographical position (EN ISO, 2019). Gazetteers play an important role in public services and geospatial data analysis. Users of location-based services often consult a gazetteer to look-up the location of administrative units, streets, addresses, etc. This is known as geocoding. Conversely, finding a description of a location on the basis of a given set of geographic coordinates, is called reverse geocoding. Because these operations are commonly performed, gazetteers are common building blocks that are used in many specific location-based services in both the public and the private sector.

Public and private sector services already rely on a number of existing gazetteers, such as Geonames.org, OpenStreetMap, Google Maps. Depending on the product selected, gazetteers are available for free or upon the payment of a fee, rely on authoritative or non-authoritative data and have different geographic and content coverage.

Under the ISA Programme, the EULF project carried out a feasibility study into the creation of an EU gazetteer service (Pignatelli et al., 2017). Most Member States have their own national gazetteer service so, if an EU gazetteer service is to be justified, there needs to be sufficient demand for pan-European applications or sufficient added value beyond existing national gazetteers.

The ELISE Action carried out a survey[6] in conjunction with EuroGeographics[7] aimed at understanding the demand-side and supply-side perspectives related to pan-European gazetteer data and services. On the demand-side, this included identifying relevant multi-national applications and the importance of authoritativeness and openness in the data. On the supply-side, respondents were asked to identify what datasets they provided and whether they were national, authoritative, open and INSPIRE compliant.

[1] https://ec.europa.eu/isa2/actions/elise_en
[2] https://ec.europa.eu/isa2/home_en
[3] https://joinup.ec.europa.eu/community/eulf/description
[4] https://joinup.ec.europa.eu/community/are3na/description
[5] https://ec.europa.eu/archives/isa/
[6] https://joinup.ec.europa.eu/collection/elise-european-location-interoperability-solutions-e-government/document/report-eu-gazetteer-survey-analysis
[7] https://eurogeographics.org/

The conclusions of the survey were that:

- There is demand for an EU gazetteer to support multi-national applications or complement existing national gazetteers, for purposes such as emergency response, searching for datasets, news items, or tourism / cultural heritage sites, validating foreign addresses;
- "Geographical names", "Administrative units" and "Addresses" are the most relevant datasets;
- Different applications have different priorities in terms of whether the data is open or authoritative – up-to-date data is most important;
- Supply side options are possible to meet this demand but across the EU not all criteria can be met from day one;
- 25% of Member States already have datasets that are complete, open, authoritative and INSPIRE-compliant;
- Geographical names and Administrative units offer more 'open' and 'authoritative' options;
- Action from ELISE, working with key stakeholders, can help bridge this gap;
- In taking any action, the goal of an EU-wide service should be part of the plan – with as much data as possible that is complete, open, authoritative and INSPIRE compliant.

In reviewing next steps following the survey, two potential pan-European services were identified that approached the overall requirement in different ways, one from EuroGeographics and another from DG ESTAT[8]. Neither option meets the 'ideal' requirement of an EU-wide gazetteer with complete (i.e. national), open, authoritative and INSPIRE-compliant datasets.

The DG ESTAT gazetteer operates to support Commission Services requirements. It is based on OpenStreetMap[9] and authoritative data. DG ESTAT has a notion to focus on authoritative data and is in the process of transforming authoritative data from some Member States to the Nominatim data model[10].

The EuroGeographics Open European Location Services (OpenELS) project, partly funded by the Commission, explored what can be made available in terms of Open European Location Services. The project developed an OpenELS licence and worked with its members on possible services. Gazetteer services are seen as a priority. The goal will be detailed authoritative INSPIRE-based services. In the short-term, EuroGeographics is in the process of putting together a less-detailed product from participating member organisations able to contribute open data.

An alternative open geographical names service providing pan-European coverage is available from Geonames.org, a global service that combines crowdsourced and authoritative data.

The different pan-European open geographical names services can be summarised as:

- AUTHORITATIVE: EuroGeographics OpenELS
- CROWDSOURCED + AUTHORITATIVE: DG ESTAT services
- CROWDSOURCED + AUTHORITATIVE: Geonames.org

[8] https://ec.europa.eu/eurostat/
[9] https://www.openstreetmap.org/
[10] https://github.com/osm-search/Nominatim

2 EU Gazetteer evaluation project

The objective of the EU Gazetteer evaluation project was to determine the suitability of available open EU gazetteer solutions based on actual demand and evaluation of the solutions.

The project comprised four main activities:

- Support to DG ESTAT data acquisition
- Development of a cultural heritage testbed
- Identification of use cases and support to their gazetteer evaluations
- Reporting on findings, recommendations and next steps (**this technical report**)

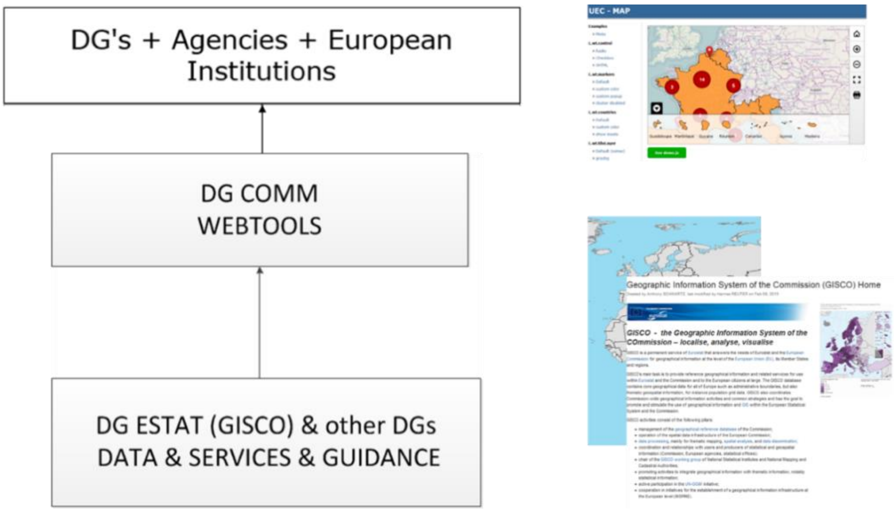
2.1 Support to DG ESTAT data acquisition

The main objective of the activity was to expand the range of authoritative gazetteer datasets available in a format compatible with the existing data used to support data and analytical requests from different Commission Services.

2.1.1 GISCO data and services

DG ESTAT through the Geographic Information System of the Commission (GISCO[11]) operates to support Commission Services requirements. GISCO is a permanent service of DG ESTAT that answers the needs of ESTAT and the European Commission for geographical information at the level of the European Union (EU), its Member States and regions. Access to data and services is possible through DG COMM Web tools (**Figure 1**) or directly from GISCO services.

Figure 1. Access to GISCO data and services



Source: Eurostat, 2019.

Some of the Corporate Level Services that are available are:

- Background Web Map Services (as background maps)
- (Reverse) Geocoding Services (i.e. find a place on a map from the description of a location, and from coordinates to a description of a location)
- Routing (i.e. how to get from one location to another by car, bicycle or lorry)
- ID – Services (i.e. find a NUTS-Code for a given location)

[11] https://ec.europa.eu/eurostat/web/gisco

GISCO Background Services use either maps based on EuroGeographics datasets or a customised OpenStreetMap (OSM) background map that is in line with the official position of the EU. This is achieved by thorough mapping of disputed areas (**Figure 2**).

Figure 2. Mapping of disputed areas

Source: Eurostat, 2019.

2.1.2 INSPIRE

The legal framework for the Infrastructure for Spatial information in the European Community (INSPIRE) has been set by the Directive (2007/2/EC) and interdependent legal acts, which are called implementing rules, in the form of Commission regulations and decisions (Cetl, et al. 2019). By design, the infrastructure itself is built upon the Spatial Data Infrastructures (SDIs) established and operated by European Union Member States that are then made compliant with the implementing rules, covering its core components: metadata, network services, interoperability of spatial datasets and services, data-sharing and monitoring and reporting, together with the obligation to establish a national coordination body.

The implementing rules for metadata, the interoperability of data themes, the network services (that help to share the infrastructure's content online) and the data sharing are complemented by non-legally binding technical guidance documents. These guidelines explain a possible technical approach for implementing the legal requirements and embed additional recommendations that may help data providers in their implementation for a range of use cases.

The thematic scope of INSPIRE includes 34 cross-sectoral categories, named themes (**Figure 3**), listed in the three annexes of the Directive and reflecting two main types of data: spatial reference data (presented in Annex I and partly in Annex II), which define a location reference that the remaining themes (in Annex III and partly in Annex II) can then refer to.

Figure 3. INSPIRE Data themes

Source: JRC, 2019.

Source data themes for establishing a gazetteer are Addresses and Geographical names, both part of Annex I.

The EU entry point to the INSPIRE infrastructure is the INSPIRE Geoportal (**Figure 4**). It serves as a central access point to the data and services from public organisations in the EU MS and EFTA countries which fall under the scope of INSPIRE.

Figure 4. Landing page of the INSPIRE Geoportal

Source: JRC, 2020.

The Geoportal consists of 2 main applications:

1. *Priority datasets Viewer*, that displays the availability and provides access to the priority datasets used for environmental reporting[12];
2. *INSPIRE Thematic Viewer*, that displays the availability and provides access to all EU MS and EFTA countries datasets falling under the scope of the INSPIRE Directive, filtered by data themes and/or countries.

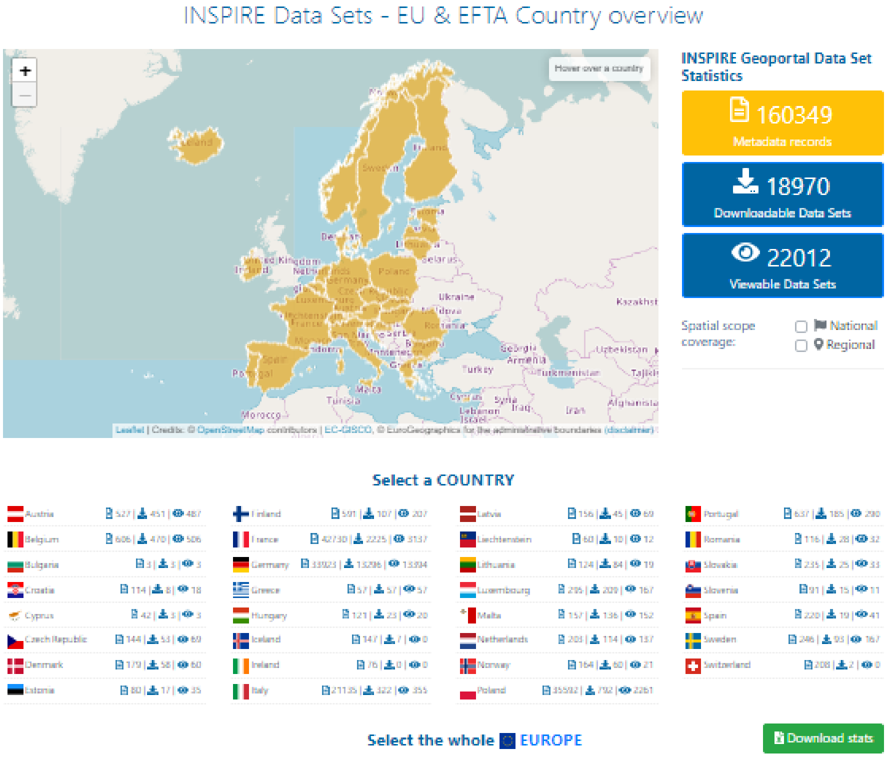
In addition, the landing page of the Geoportal provides a link to access the *INSPIRE Reference Validator*, that helps data providers check whether their datasets, services and metadata meet the INSPIRE requirements.

The INSPIRE Geoportal enables cross-border data discovery, visualisation and download. The Geoportal does not store any geospatial data, but it simply acts as the main client application of the whole INSPIRE infrastructure by exposing data through the harvesting of the CSW endpoints made available by MS and EFTA countries (currently 36 Discovery Services from 32 countries).

[12] https://ies-svn.jrc.ec.europa.eu/projects/2016-5/wiki

Insight into the current status of the infrastructure is provided by the INSPIRE Thematic Viewer, which offers two possibilities for browsing datasets: by individual EU MS & EFTA country and by INSPIRE data theme. **Figure 5** shows the availability of datasets in EU MS and EFTA countries as of 15/05/2020.

Figure 5. Availability of INSPIRE datasets in EU MS and EFTA countries on 15/05/2020



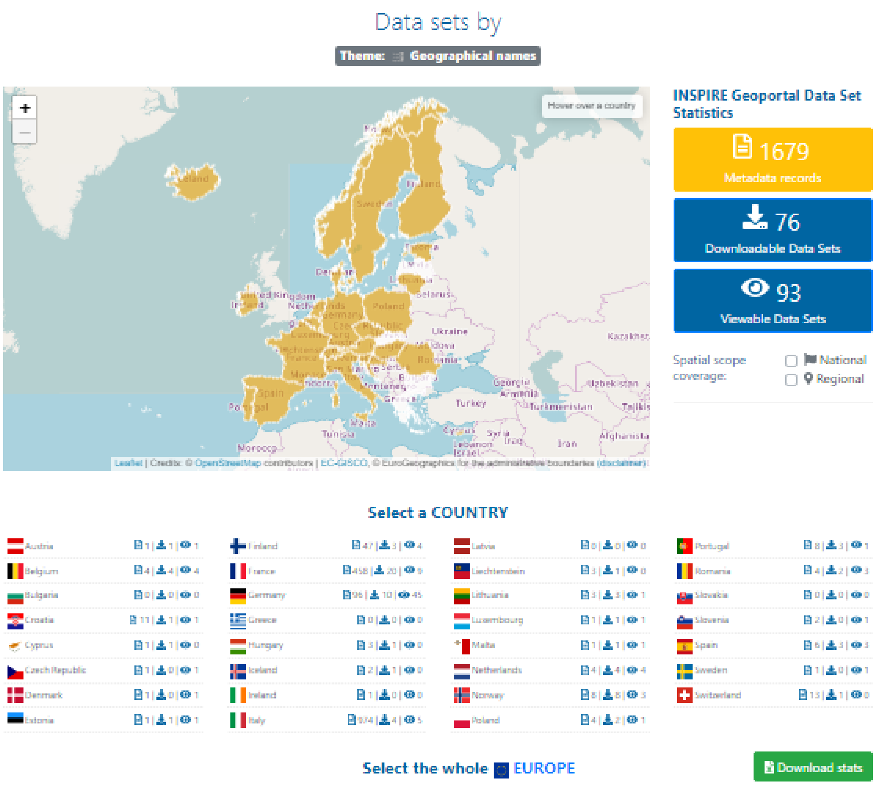
Source: JRC, 2020.

The three numbers related to each country correspond to the number of available metadata records, downloadable datasets and viewable datasets. Metadata can be filtered by country and data theme. For example, **Figure 6** shows Addresses (AD) and **Figure 7** shows Geographical Names (GN).

Figure 6. Availability of Addresses (AD) on 15/05/2020

Source: JRC, 2020.

Figure 7. Availability of Geographical Names (GN) on 15/05/2020

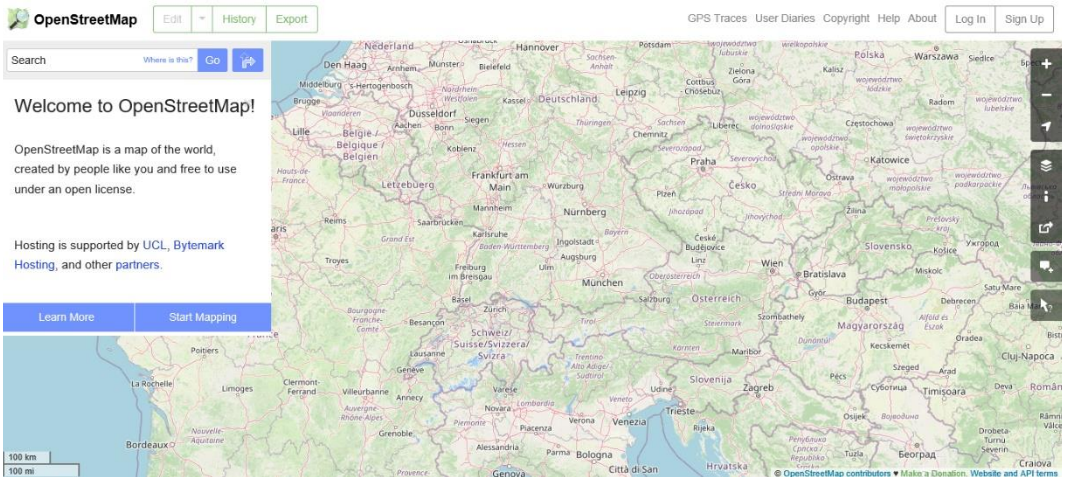


Source: JRC, 2020.

2.1.3 OpenStreetMap (OSM)

OpenStreetMap (OSM) is a collaborative project to create an editable and free geospatial database of the whole world[13] (**Figure 8**). Although OSM itself is a database, the main output generated from the database is the map available on the main website[14].

Figure 8. OpenStreetMap map



Source: © OpenStreetMap contributors, 2020.

OSM data are created from geographic information captured with mobile GPS devices, orthophotographs and other sources using specific editing tools created by the OSM community[15]. Both the OSM database and the derived cartography (tiles) are distributed under the open access Open Database License (ODbL).

OSM uses a topological data structure. The data is stored in the Mercator projection datum WGS84 lat/lon (EPSG:4326). The primitive data or basic elements of the OSM cartography are:

- Nodes. They are points that collect a given geographical position.
- Ways. They are an ordered list of between 2 and 2000 nodes representing both polylines and polygons.
- Relations. They are used for polylines or polygons of more than 2000 nodes as well as for groups of nodes, ways and/or other relations to which certain common properties can be assigned: for example, all those ways that are part of the Way of St. James.
- Tags. They can be assigned to nodes, ways or relations and consist of a key and a value. For example: highway=trunk defines a road as a trunk road. Each OSM object must have at least one tag, but there is no limit to the number of tags that an object can have.

Based on this structure of data, processes have been created to extract the Addresses data. In OSM, the tags defining Addresses can be associated to nodes, ways and relations.

2.1.4 Data acquisition and transformation

The goals of data acquisition were to:

- expand the range of authoritative gazetteer datasets available to GISCO in a format compatible with the existing data they use to support data and analytical requests from different Commission Services; and
- determine how to best use the new data available to them.

[13] https://wiki.openstreetmap.org/wiki/Main_Page
[14] https://www.openstreetmap.org
[15] https://wiki.openstreetmap.org/wiki/Editors

The first task was Data transformation that included:

1. Identification of the relevant Addresses data to be extracted from the source datasets (see below)
2. Extraction of subset datasets (including only the required information) from the source datasets
3. Transformation of the extracted datasets to the target (Nominatim) OSM format
4. Loading transformed datasets into the Nominatim database
5. Resolving any errors in the extract, and load processes

The Addresses data sources to be incorporated to the Nominatim database were for the following countries: BE, NL, LU, FR, ES, CZ. In the case of BE the comparison was limited to NUTS 1 BE100 since only postal addresses for this NUTS level were available and data for the whole national territory were not available.

An address is location of properties based on address identifiers, usually by road name, house number, postal code[16]. In INSPIRE model, an address is an identification of the fixed location of a property. The full address is a hierarchy consisting of components such as geographic names, with an increasing level of detail, e.g.: town, then street name, then house number or name. It may also include a post code or other postal descriptors. The address may include a path of access, but this depends on the function of the address.

The preliminary analysis of the input data sources was performed by looking for data available through the INSPIRE Geoportal[17] as a main data source. The objective was to find and download Addresses authoritative data. Data sources were obtained from the authoritative official publications of their respective countries. The OSM data were already available in the GISCO database.

Country reports

Table 1. Belgium

Country	Belgium (BE)
Nº of files	1 (220.382 registers)
File format	shp
Coverage	

[16] https://inspire.ec.europa.eu/theme/ad
[17] https://inspire-geoportal.ec.europa.eu/

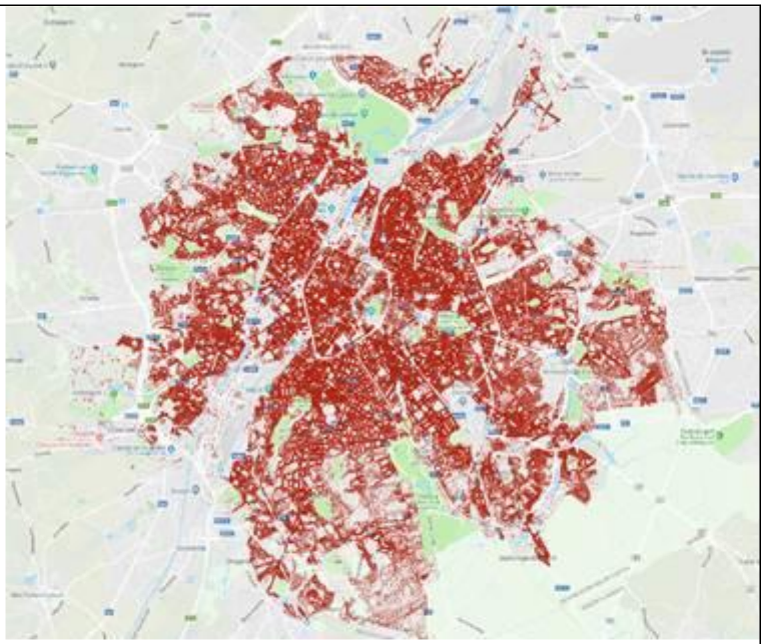
	

Table 2. Czechia

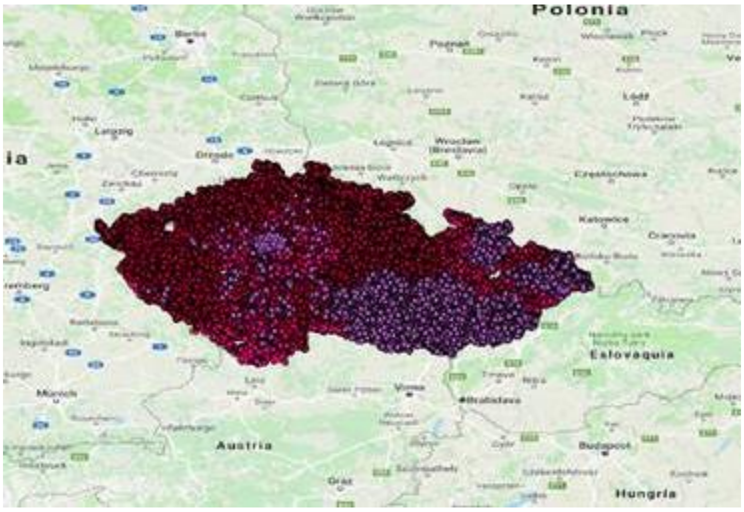
Country	Czechia (CZ)
Nº of files	6258 files. 2866154 registers in total.
File format	xml
Coverage	

Table 3. Spain

Country	Spain (ES)
Nº of files	7604 files (619 with no data) (14.163.269 registers)

File format	gml
Coverage	

Table 4. France

Country	France (FR)
Nº of files	102 files (28 408 053 registers)
File format	csv
Coverage	

Table 5. Luxembourg

Country	Luxembourg (LU)
Nº of files	1 (166. 539 registers)
File format	shp

Coverage	

Table 6. Netherlands

Country	Netherlands (NL)
Nº of files	1 (9 103 995 registers)
File format	csv
Coverage	

The data transformation flow included several subtasks to transform the original datasets to OSM format files (**Figure 9**).

Figure 9. Data transformation flow

Source: JRC, 2020.

Starting from the datasets originally provided, the analysis and design of the input data models was carried out in order to harmonise these datasets with the target data model (OSM format file) for uploading into the Nominatim database.

The transformation data flow included the following subtasks to transform the original datasets to OSM format files:

- An initial analysis was performed to determine a rough estimation status for the quality of the data sources provided and their suitability to be used as input data for the Nominatim database. As a result of this activity, a preliminary assessment of the quality for the provided data was done.

- The datasets were provided in many formats, and they needed to be explored in order to determine the correspondence of the source data to the OSM data model. GIS tools as QGIS[18] were used as desktop solutions for the manual inspection of the datasets.
- The dataset coverage, projection and layers were reviewed at this level to identify the correspondence of the required elements (Addresses features).
- Extraction of subset datasets (including only the required information) from the source dataset. The identified elements were extracted in a dedicated ETL process because of the different data source formats.
- The process of extraction of the elements, reprojection (if necessary) and transformation to OSM geometry elements was performed.
- Transformation of the specified datasets to the target OSM format was performed.
- Creation of an OSM XML file with geometry and attributes of each of the elements was performed so they can be exported into a Nominatim database. The elements in the OSM XML file were labelled at this stage so they can be compared later on with original data.
- OSM XML files were loaded into the Nominatim database.

For the ETL (Extract Transform Load) processes the GeoKettle[19] software was used.

2.1.5 Data comparison

Analysis of data sources

Table 7 shows the total numbers of Addresses per country from both the authoritative data sources and OSM.

Table 7. Total Addresses

DATA SOURCE	TOTAL ADDRESSES
authoritative_be	220.382
authoritative_cz	2.866.153
authoritative_es	14.127.481
authoritative_fr	28.408.044
authoritative_lu	166.539
authoritative_nl	8.805.088
osm_be (BE100)	221.652
osm_cz	1.538.324
osm_es	971.649
osm_fr	2.923.973
osm_lu	57.700
osm_nl	9.214.685
total_authoritative	**54.593.687**
total_osm	**14.927.983**

[18] https://qgis.org
[19] https://live.osgeo.org/archive/10.0/en/overview/geokettle_overview.html

A series of SQL statements were executed to test the quality of the alphanumeric and spatial components of the source information:

- Quality control of the alphanumeric component
- Percentage of records by source in relation to the total for the country
- Number of records without "street name"
- Number of records without "house number"
- Number of records without "zip code"
- Number of records without "locality" or "city"
- Number of records without "municipality"
- Number of records without "country"
- Quality control of the spatial component
- Postal addresses outside the boundaries of the country of origin
- Number of postal addresses by NUTS level
- Quality control of coherence between the alphanumeric component and the spatial component reported in the datasets
- Same geometry and same alphanumeric address (duplicate records).
- Same geometry with different alphanumeric address.

Comparisons between the two data sources

The following comparisons were made between the two data sources (authoritative and OSM Addresses):

- Number of records having the same location (geometry) and street name in the authoritative dataset and in the OSM dataset
- Number of records having the same location (geometry) and different street name in the authoritative dataset and in the OSM dataset
- Number of records having different location (geometry) and the same street name and house number in the authoritative dataset and OSM dataset
- Number of records that have the same location (geometry) and different street name and different house number in the authoritative dataset and OSM dataset.

Comparisons with other reference sources

Quality control against other reference information sources (Postal codes and NUTS):

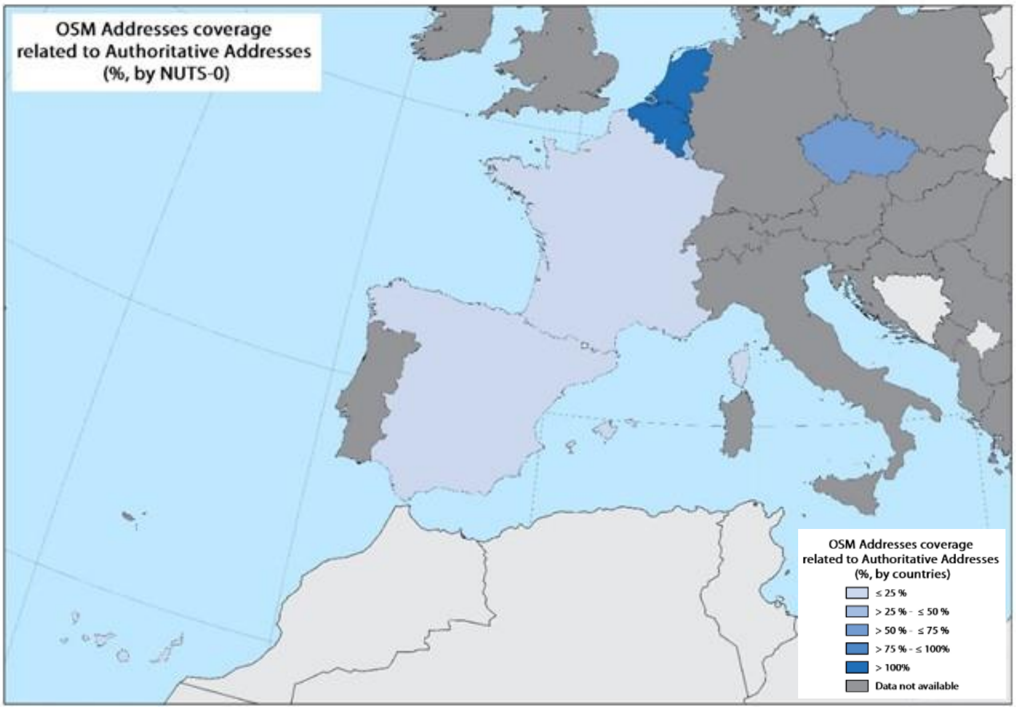
- Number of records matching Postal Code with PostalCode.csv (Postal Code, NUTS3 Code)
- Number of records matching Postal Code with PostalCode.csv (Postal Code, NUTS3 Code) and located in the NUTS corresponding to the alphanumeric postal code
- Number of records by NUTS levels.

Results

Cartographic and/or tabulated results of quality tests and comparisons for the most significant tests are presented below.

OSM Addresses coverage related to Authoritative Addresses by NUTS-0 level is shown in **Figure 10**. The percentage of Addresses from authoritative sources is calculated in relation to the total obtained (OSM/Authoritative ratio) for the NUTS-0 level (countries) by making a spatial join between the layers of authorities and OSM with the NUTS layer from GISCO. This spatial join eliminates from the percentage those addresses that are located outside the limits of their corresponding country.

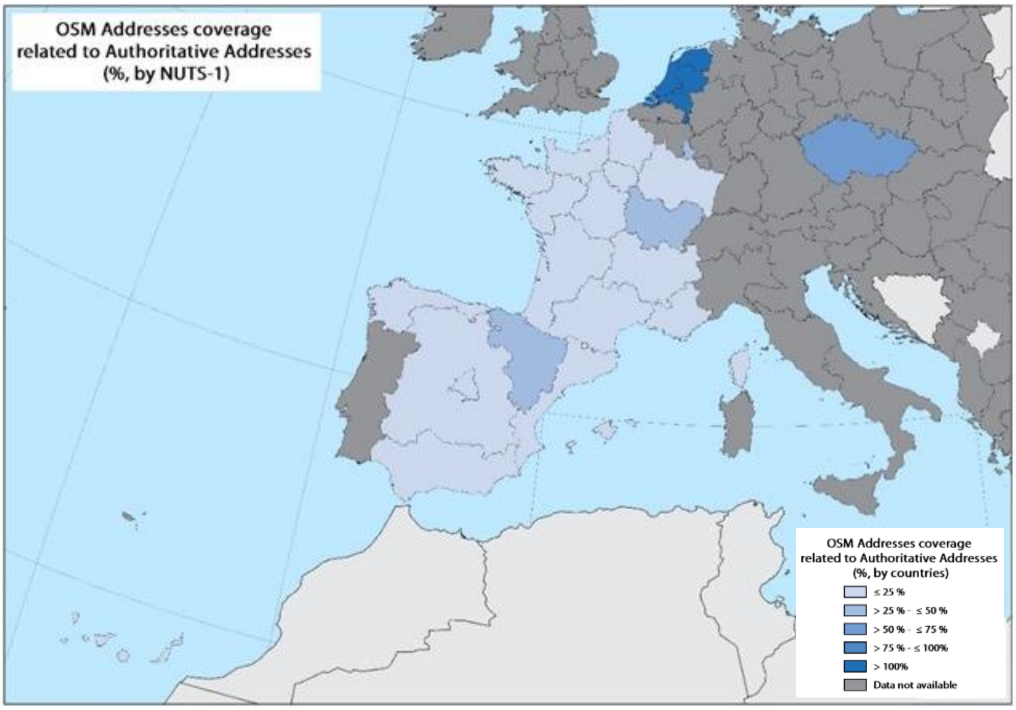
Figure 10. OSM Addresses coverage related to Authoritative Addresses by NUTS-0 level



Source: JRC, 2020.

OSM Addresses coverage related to Authoritative Addresses by NUTS-1 level (**Figure 11**) is calculated in relation to the total obtained (OSM/Authoritative ratio) for NUTS-1 level by making a spatial join between the layers of authorities and OSM with the NUTS layer from GISCO. This spatial join eliminates from the percentage those addresses that are located outside the limits of their corresponding country.

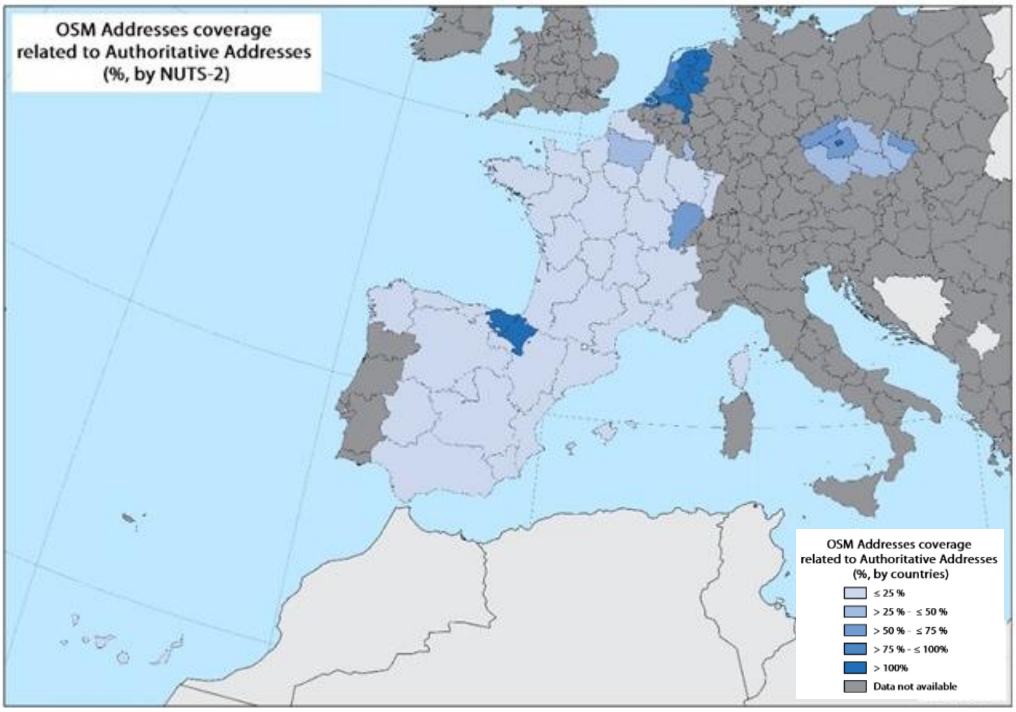
Figure 11. OSM Addresses coverage related to Authoritative Addresses by NUTS-1 level



Source: JRC, 2020.

OSM Addresses coverage related to Authoritative Addresses by NUTS-2 level (**Figure 12**). The percentage of addresses from authorities is calculated in relation to the total obtained (OSM/Authoritative ratio) for NUTS-2 level by making a spatial join between the layers of authorities and OSM with the NUTS layer from GISCO. This spatial join eliminates from the percentage those addresses that are located outside the limits of their corresponding country.

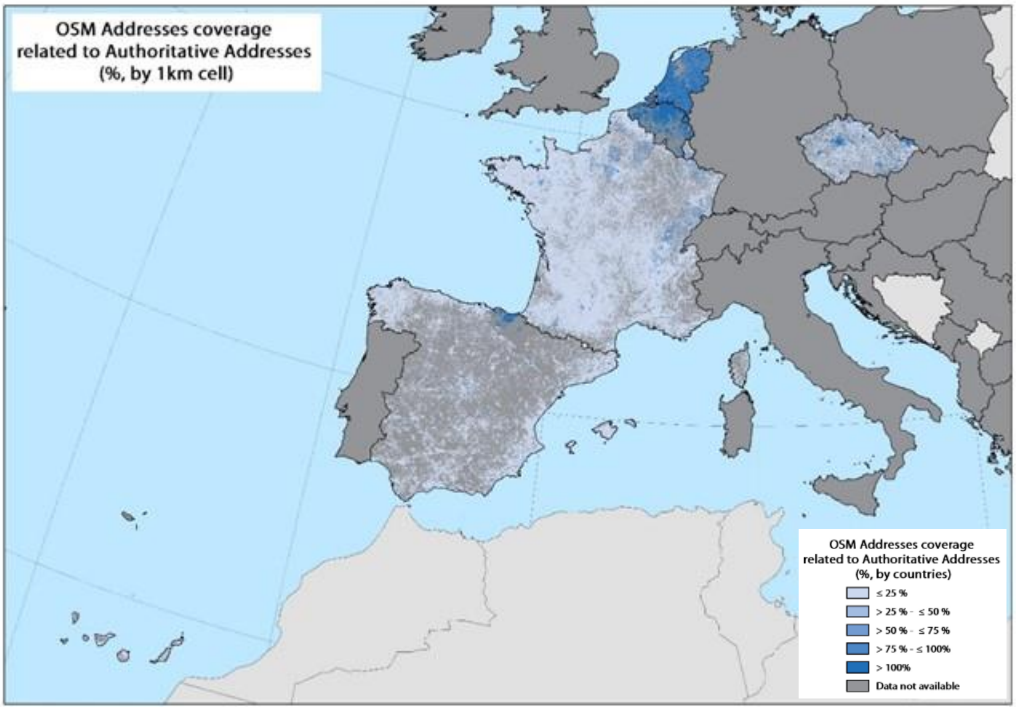
Figure 12. OSM Addresses coverage related to Authoritative Addresses by NUTS-2 level



Source: JRC, 2020.

The following map (**Figure 13**) represents the OSM Addresses coverage related to Authoritative Addresses per 1Km cell.

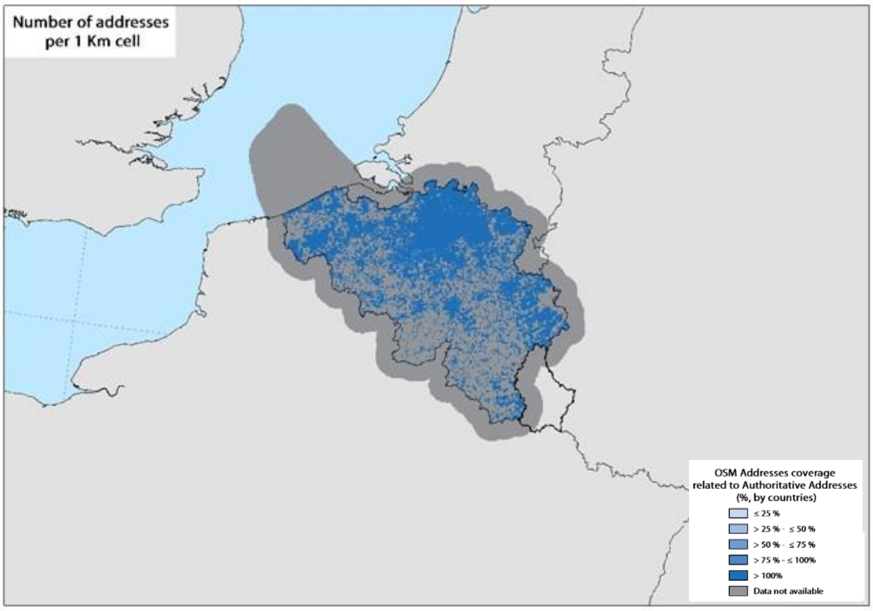
Figure 13. OSM Addresses coverage related Authoritative Addresses per 1Km cell



Source: JRC, 2020.

The following map (**Figure 14**) represents OSM Addresses coverage related to Authoritative Addresses per 1Km for BE.

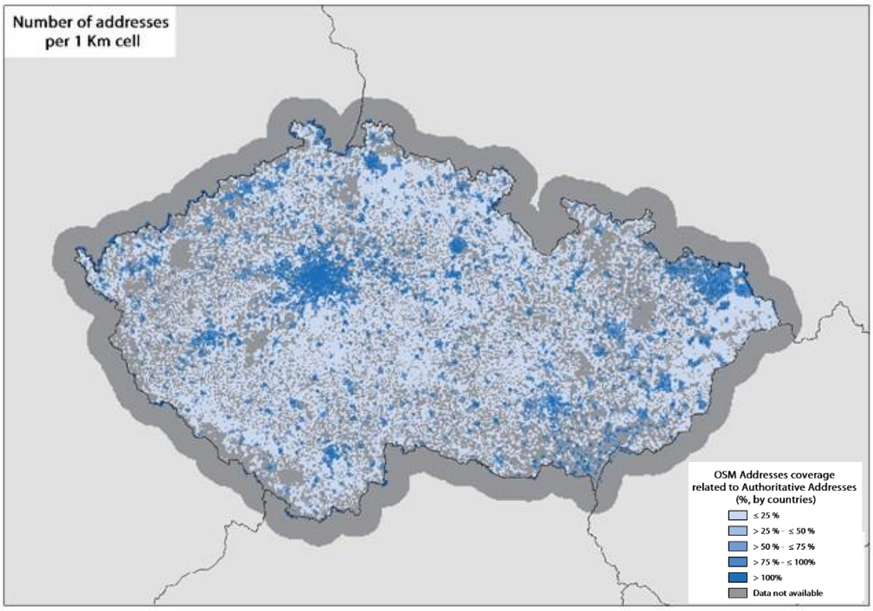
Figure 14. OSM Addresses coverage related to Authoritative Addresses per 1Km for BE



Source: JRC, 2020.

The following map (**Figure 15**) represents OSM Addresses coverage related to Authoritative Addresses per 1Km for CZ.

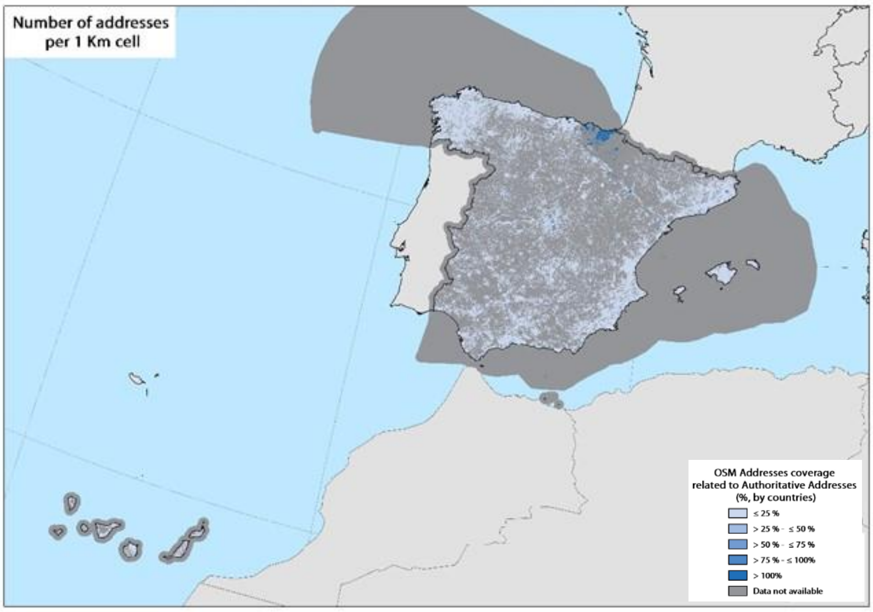
Figure 15. OSM Addresses coverage related to Authoritative Addresses per 1Km for CZ



Source: JRC, 2020.

The following map (**Figure 16**) represents OSM Addresses coverage related to Authoritative Addresses per 1Km for ES.

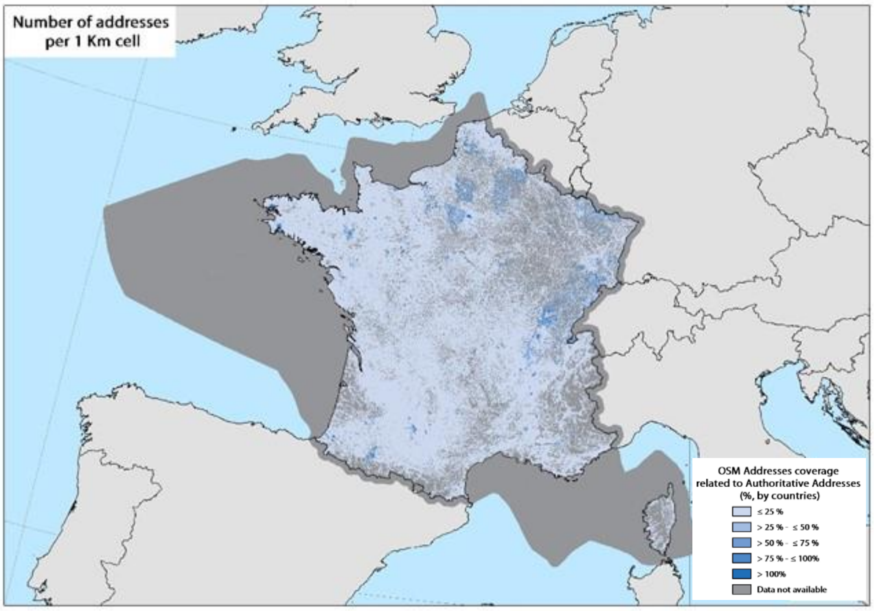
Figure 16. OSM Addresses coverage related to Authoritative Addresses per 1Km for ES



Source: JRC, 2020.

The following map (**Figure 17**) represents OSM Addresses coverage related to Authoritative Addresses per 1Km for FR.

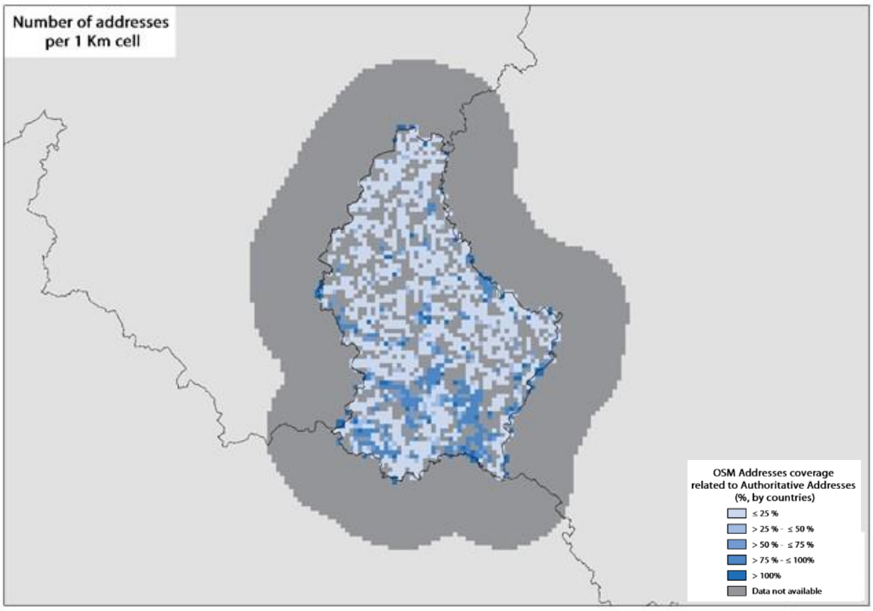
Figure 17. OSM Addresses coverage related to Authoritative Addresses per 1Km for FR



Source: JRC, 2020.

The following map (**Figure 18**) represents OSM Addresses coverage related to Authoritative Addresses per 1Km for LU.

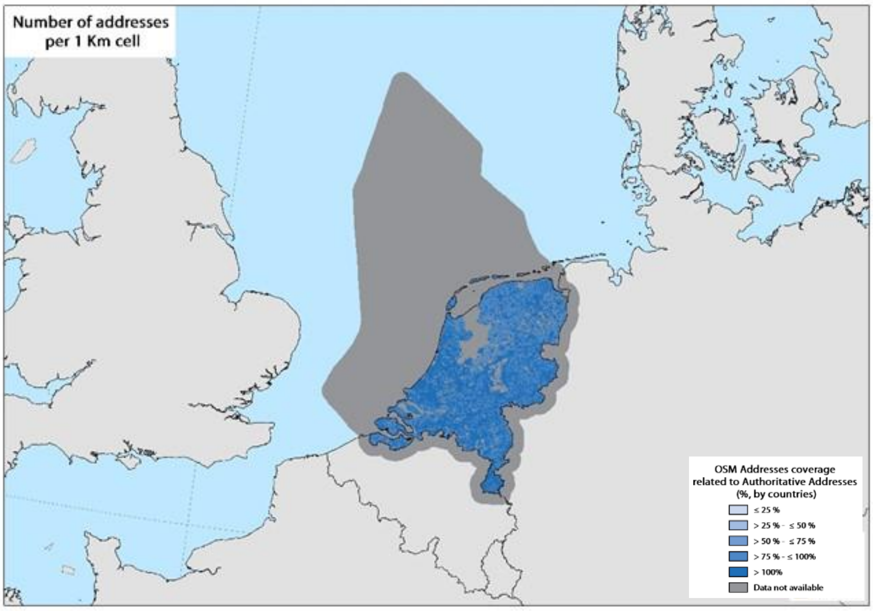
Figure 18. OSM Addresses coverage related to Authoritative Addresses per 1Km for LU



Source: JRC, 2020.

The following map (**Figure 19**) represents OSM Addresses coverage related to Authoritative Addresses per 1Km for NL.

Figure 19. OSM Addresses coverage related to Authoritative Addresses per 1Km for NL



Source: JRC, 2020.

Table 8 shows comparisons between the Addresses datasets from the two data sources (authoritative and OSM).

Table 8. Comparison between authoritative and OSM Address datasets

AUTHORITATIVE VS OSM	SAME LOCATION AND SAME STREETNAME	SAME LOCATION AND DIFFERENT STREETNAME	DIFFERENT LOCATION AND SAME STREETNAME AND SAME HOUSENUMBER	SAME LOCATION AND DIFFERENT STREETNAME AND DIFFERENT HOUSENUMBER
authoritative_osm_BE	2	0	179,935	0
authoritative_osm_CZ	1,857	5	2,806,708	0
authoritative_osm_ES	0	329	191	3
authoritative_osm_FR	0	227	0	0
authoritative_osm_LU	45	5	170,296	0
authoritative_osm_NL	703,261	0	179,935	0

Table 9 shows quality control of the authoritative and OSM Address datasets against other reference information sources (Postal codes and NUTS).

Table 9. Number of records matching Postal Code (Postal Code, NUTS3 Code)

COUNTRY	# AUTHORITATIVE PC MATCHED	% AUTHORITATIVE PC MATCHED	# OSM PC MATCHED	% OSM PC MATCHED
BE	220.381	78,20 %	61.425	21,80 %
CZ	2.571.658	99,97 %	717	0,03 %
ES	11.133.328	94,16 %	690.107	5,84 %
FR	25.799.154	97,10 %	769.520	2,90 %
LU	164.899	76,67 %	50.176	23,33 %
NL	8.802.708	49,28 %	9.058.463	50,72 %

The percentage of correspondence by data source in relation to total correspondence by country is shown above for each source country.

As main results of the analysis, the following conclusions can be drawn:

- OSM provides higher completeness in Addresses for large population areas.
- Authoritative Addresses provide better results in rural areas. This is in line with findings in the literature, which show that highly populated, urban areas are usually well mapped than poorly populated, rural area where the OSM community may be less active.
- It is necessary to mention the case of NL, where the authoritative dataset was imported into the OSM database and has been maintained by the community since then. In this case, OSM Addresses are comparable to Authoritative Addresses throughout the country.
- In general, the main conclusion is that both data sources are complementary and mutually enhancing results can be obtained by combining the two.

2.1.6 Data switching

The objective of data switching was to provide ESTAT's Nominatim services with the ability to deliver results from different sources in a clustered manner. In this way, users can benefit from the combination and accuracy of each of the combined Nominatim instances.

Data switching allows:

- Authoritative and OpenStreetMap Addresses can be used complementary
- Mechanism to share and offer data enables support for the following:
  - Multiple data sources
  - Common interface
  - Selecting and switching between data of different provenance
  - Minimizing impact in existing applications

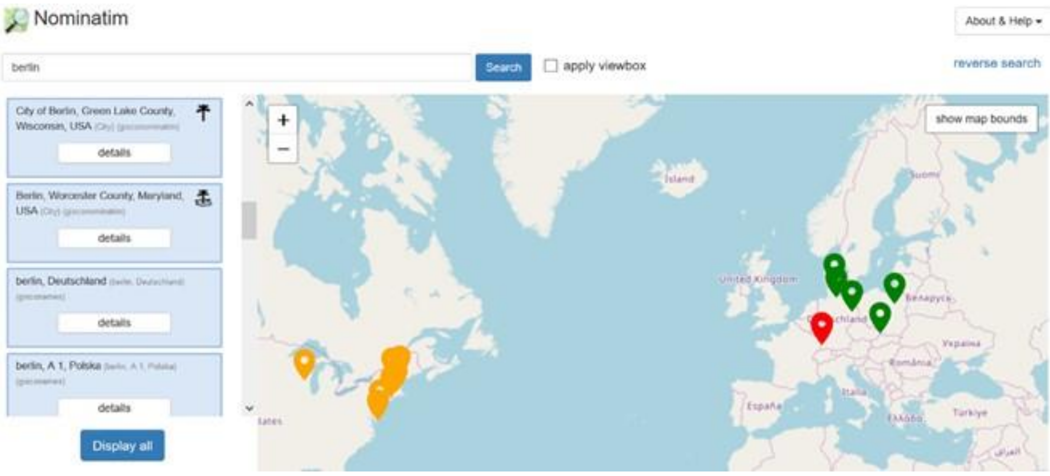
Different approaches to the problem were analysed, resulting in the development of a proxy application that integrates the response of different Nominatim instances. Two interfaces were developed for search (by name) and reverse search (by coordinates).

Interfaces were developed for both searches and reverse searches (**Figure 20** and **Figure 21**) based on the following output formats:

- json

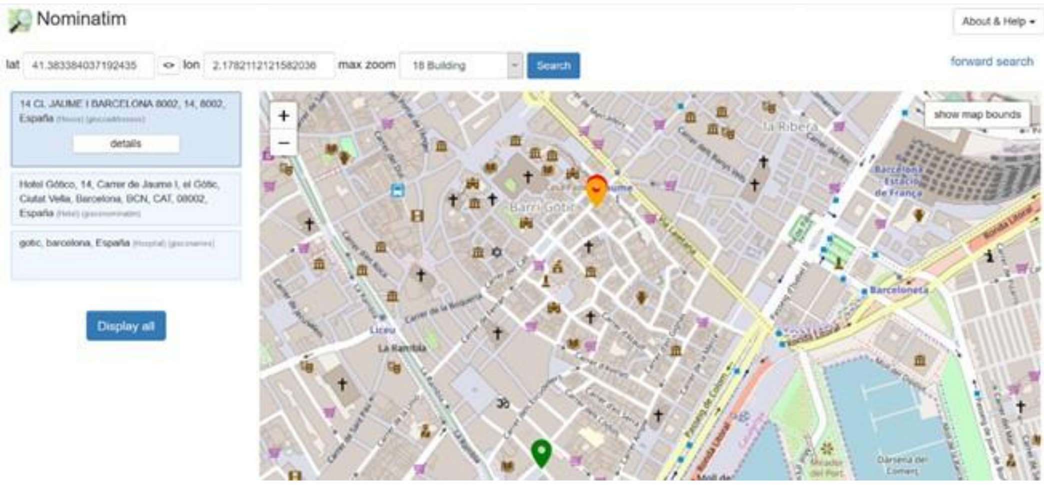
- jsonv2
- geojson
- geocodejson
- xml
- html

Figure 20. Search interface



Source: JRC, 2020.

Figure 21. Reverse search interface



Source: JRC, 2020.

2.2 Development of a cultural heritage testbed

The objective of the "cultural heritage testbed" application was to identify data, functionality gaps and improvements needed in different gazetteer solutions, namely EuroGeographics, DG ESTAT and Geonames.org, when comparing against typical 'cultural heritage' requirements. The focus was on the INSPIRE data theme Geographical Names (GN)[20].

"Cultural heritage is seen as 'the total ways of living built up by a group of human beings, which is passed from one generation to the next', given to them by reason of their birth. Geographical names are addresses, the keys to identifying specific places, but also of irreplaceable cultural value of fundamental importance to local identity, and a person's sense of belonging, and therefore must be protected and preserved"[21]. In this context GN are seen as cultural heritage, thus the naming 'cultural heritage testbed'.

A geographical name serves as a means to identify a location. Gazetteers and gazetteer services associate the names with corresponding features – or locations – by means of coordinates, feature types and/or other necessary information. GN or toponyms are names of areas, regions, localities, cities, suburbs, towns or settlements, or any geographical or topographical feature of public or historical interest. A geographical name is a proper noun applied to a natural, man-made or cultural feature on Earth. A feature can have different names in one or several languages and the names may be provided, together with appropriate information on the feature, in different products like maps and gazetteers as well as respective services.

An endonym is a name for a geographical feature in an official or well-established language occurring in that area where the feature is situated. An exonym is a name used in a specific language for a geographical feature situated outside the area where that language is widely spoken, and differing in form from the respective endonym(s) in the area where the geographical feature is situated (UNGEGN, 2007).

The idea behind the development of a cultural heritage testbed was to explore the language and temporal dimensions of geographical names, e.g.:

- Toponyms can be in different languages (e.g. Wien in AT)
    - Beč, Beç, Bech, Bécs, Bin (빈)), Dunaj, Fienna, Vedunia, Vena, Vídeň, Viden, Viedeň, Viên, Viena (Виена), Vienna, Vienne, Viénni, Vieno, Viin, Vina, Vínarborg, Vindobona, Vīne, Viyana, Vjenë, Vjenna, Vyana, Weiyena (維也納), Wene, Wenen, Wiedeń, Wīn (ウィーン), Wina, فيينا, وين
- The same toponyms in different languages (i.e. exonyms) may describe different objects
    - Dunaj (Slovenian) = Vienna (City), Dunaj (Slovakian) = Danube (river)
- Name changes over time

Development was divided into several tasks (**Figure 22**).

Figure 22. "cultural heritage testbed" development

Source: JRC, 2020.

The application was not intended to be a fully-functioning application suitable for moving into production, but rather to provide information about the results obtained from the different providers using their APIs as each of the gazetteers has a different service:

- EuroGeographics[22]

[20] https://inspire.ec.europa.eu/theme/gn
[21] https://unstats.un.org/unsd/geoinfo/ungegn/docs/Bulletin/UNGEGN_Bulletin_48_FINAL.pdf
[22] https://openels.eu/products/

- DG ESTAT[23]
- Geonames.org[24]

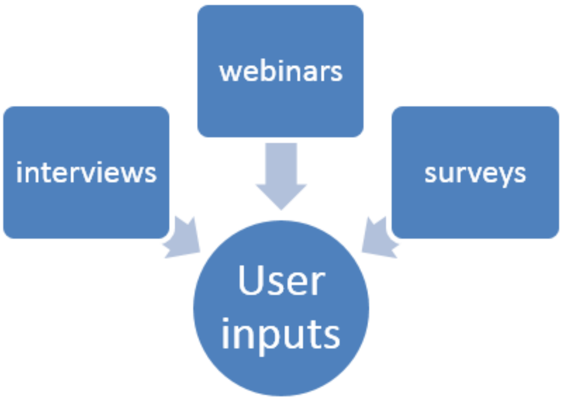
2.2.1 User Requirements

Requirements were collected from an expert group of users in order to build the testbed application so it could provide them with significant results. The aim was to identify detailed requirements from potential users in order to determine what needs could or could not be met by available gazetteer services. Multiple inputs were used in order to gather all the details related to temporal and multilingual requirements as well as to the use of persistent identifiers.

The approach for this task consisted in the use of different tools and techniques to gather requirements from the users (**Figure 23**), as follows:

- Webinars, where the relevant information and the main approach were shared with the users so a common understanding could be obtained.
- Survey, as a collection of relevant questions in order to target further activities.
- Interviews, virtual meetings with the users to exchange impressions and needs for the testbed application.

Figure 23. Gathering of user requirements



Source: JRC, 2020.

The main outcomes were:

- Interviews
  - Design and functionality discussions
- Webinars
  - Short-term plan for design/requirement group was presented
  - First design of the Testbed application was shared with the attendants
  - Final requirements analysis for the testbed refining the collected information was performed
- Surveys. A questionnaire was distributed in order to collect requirements in a systematic way:
  - Testbed functionality / user experience was retrieved
  - Search term list / example datasets were obtained to help with the development

These requirements are summarised in the following table.

[23] https://nominatim.org/release-docs/latest/api/Overview/
[24] http://www.geonames.org/export/web-services.html

Table 10. User requirements

US #	AS A…	I WANT TO…	SO THAT…
1	user	select gazetteers	I'm not forced to use them all
2	user	search a location by its name	I can compare gazetteers
3	user	search **multiple locations** by uploading a file	I can run a batch request
4	user	see the results on a map	I can validate the real location
5	user	see results in tabular format	I can easily compare the details
6	user	**score** the gazetteers based on distance/semantic accuracy	I can easily know which result is best
7	user	download the results in a csv format	I can work on a deeper analysis

This was the basis for the data model development.

2.2.2 Data model

The target data model was based on the user requirements collected during the webinars and with the questionnaires. The main requirements for the Cultural Heritage Testbed were the following:

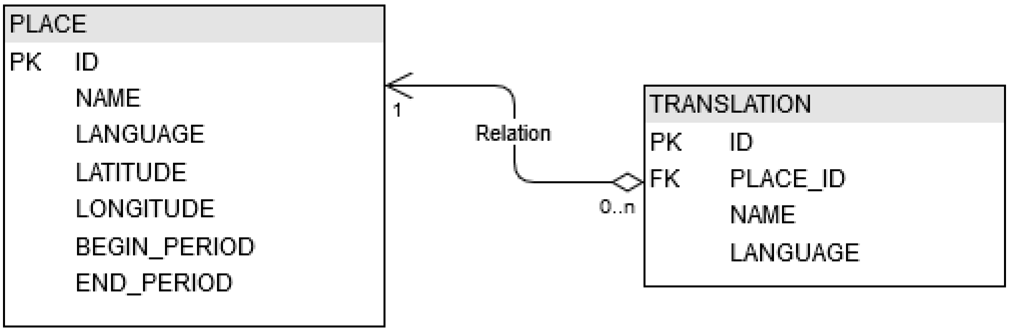
- Search for a place based on its name and obtain its location
- Compare the results of different gazetteers
- Use different languages for the search
- Search by temporal dimension.

These requirements were mapped in according to the following concepts:

- Each PLACE must have a permanent ID
- Each PLACE must have a PLACE NAME
- Each PLACE must have a LANGUAGE
- Each PLACE must have a LATITUDE and a LONGITUDE
- Each PLACE must have a BEGIN PERIOD and an END PERIOD
- Each PLACE may be related with 0 or more TRANSLATIONs
- Each TRANSLATION must have an ID
- Each TRANSLATION must have a NAME
- Each TRANSLATION must have a LANGUAGE

Figure 24 below shows the target conceptual data model.

Figure 24. Target conceptual data model



Source: JRC, 2020.

2.2.3 Gap analysis between target and existing data models

During the development, different gazetteer services were used but with no access to the existing underlying data models. In this regard, only the different responses provided by the services were analysed to check if the services satisfy the following main requirements for the Cultural Heritage Testbed:

- Search for a place based on its name and obtain its location.
- Compare the results of different gazetteers.
- Use different languages for the search.
- Search by temporal dimension.

Services analysed were:

- EC - GeoNames, using Nominatim interface
  - https://europa.eu/webtools/rest/gisco/names/search.php?&format=json&limit=1&q=Vitoria-Gasteiz
- EC - OSM, using Nominatim interface
  - https://europa.eu/webtools/rest/gisco/nominatim/search.php?&format=json&limit=1&q=Vitoria-Gasteiz
- EuroGeographics
  - https://www.euro-geo-opendata.eu/api/v2/maps/external/wfs/open-regional-gazetteer-service?SERVICE=WFS&request=GetFeature&VERSION=2.0.0&token=...&typename=gn:NamedPlace&srsname=EPSG:4326&count=1&FILTER=...
- GeoNames.org
  - http://api.geonames.org/searchJSON?username=p2geo&maxRows=1&q=Vitoria-Gasteiz

The services all support searching places in different languages. The main gap is with the temporal aspects. None of the gazetteers include the option to search in a specific time period.

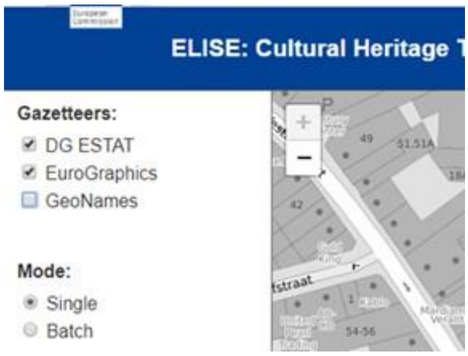
Additionally, in relation to the temporal search capabilities, the data should be enriched with historical names in order to make this functionality useful.

2.2.4 "Cultural Heritage Testbed" Application

The development of the Testbed application was achieved through an iterative incremental approach, offering users developments of the application as the requirements were being defined. In this sense, the approach was successful, as it has allowed iteration and definition in more detail and clearer feedback on the application as a whole.

[US #1] Functionality: gazetteers selection

All gazetteers are selected by default:



- DG ESTAT GeoNames (GN)
- DG ESTAT OpenStreetMap service (OSM)

CROWDSOURCED + AUTHORITATIVE

- EuroGeographics OpenELS

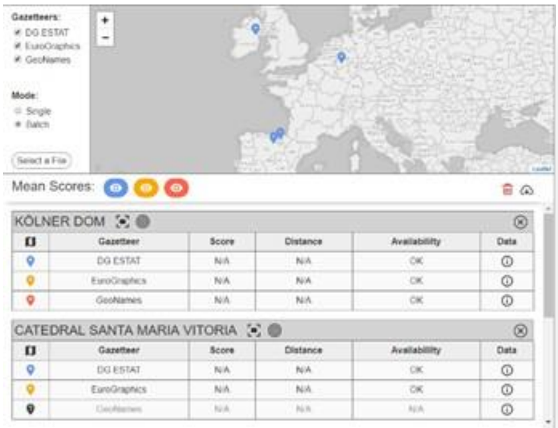
AUTHORITATIVE

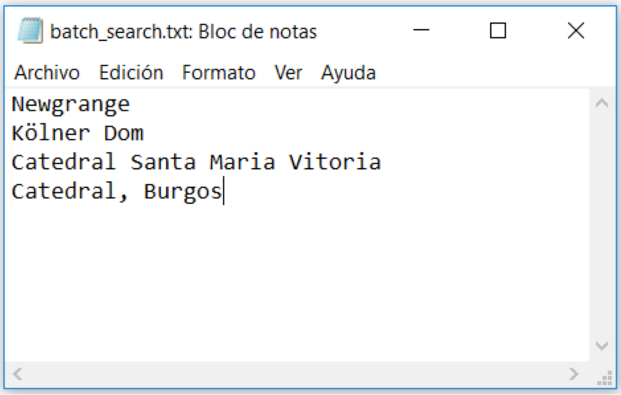
- Geonames.org

CROWDSOURCED + AUTHORITATIVE

[US #2] Functionality: single search

[US #3] Functionality: batch search





[US #4] Functionality: results – map viewer

[US #5] Functionality: results – table view

KÖLNER DOM					
	Gazetteer	Score	Distance	Availabililty	Data
	DG ESTAT	9.9	1m	OK	ⓘ
	EuroGraphics	9.9	1m	OK	ⓘ
	GeoNames	9.3	14m	OK	ⓘ

[US #6] Functionality: distance scoring (distance calculated between the searched coordinates and the original coordinates from user's request file or target location setting)

[US #7] Functionality: download results

All results displayed in tables can be downloaded in csv format

NEWGRANGE					
	Gazetteer	Score	Distance	Availabililty	Data
	DG ESTAT	9.9	[illegible]	OK	ⓘ
	GeoNames	1.6	[illegible]	OK	ⓘ

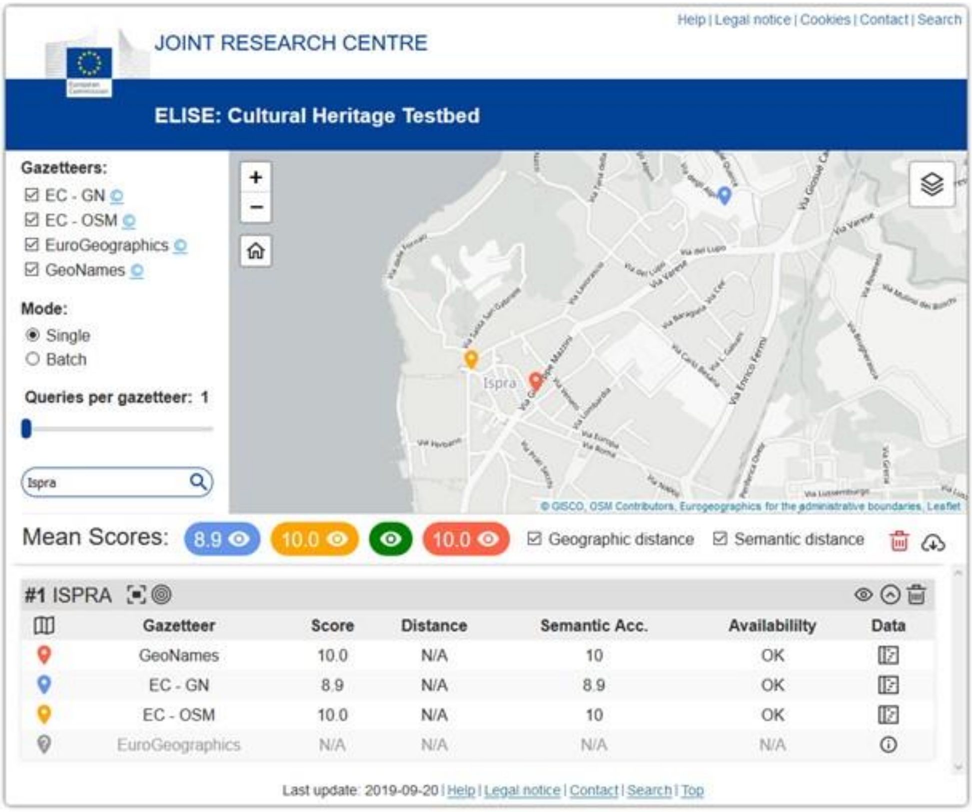
KOLN CATHEDRAL					
	Gazetteer	Score	Distance	Availabililty	Data
	DG ESTAT	9.6	8m	OK	ⓘ
	GeoNames	9.5	10m	OK	ⓘ

CATEDRAL BURGOS

	A	B	C	D	E	F	G	H
1	address	gazetteer	performance	availability	distance	score	latitude	longitude
2	Catedral Bur	dgestat	404ms	OK	304,58677	3,93579958	42,3387302	-3,7019779
3	Catedral Bur	geonames	1775ms	N/A			none	none
4	Koln Cathedı	dgestat	552ms	OK	12,410485	9,36427405	50,9413395	6,95813309
5	Koln Cathedı	geonames	2026ms	OK	13,7555993	9,30032649	50,94123	6,95823
6	Newgrange	dgestat	390ms	OK	1,06147448	9,94217713	53,6946974	-6,47555217
7	Newgrange	geonames	2472ms	OK	1393,79726	1,63052419	53,68333	-6,46667
8								

The application (**Figure 25**) enables search and comparison of results according to the users requirements, offering a comparison interface to the users.

Figure 25. "Cultural Heritage Testbed" Application



Source: JRC, 2020.

Using the "cultural heritage testbed" application, it is possible to evaluate which EU Gazetteer solution is the most suitable to satisfy the requirements on Cultural Heritage to geolocate places from their names and obtain the name for a given coordinate.

The application source code together with the testbed help documentation is available and can be provided by JRC to interested users.

2.3 Use cases

Gazetteers play an important role in enabling location-based services, for example to look-up the location of administrative units, streets, and addresses or find a description of a location on the basis of a given set of geographic coordinates. Public and private sector services already rely on a number of existing gazetteers, such as already described in the previous sections: Geonames.org, OpenStreetMap, Google Maps.

Gazetteers are more than just lists of places. They are invaluable tools for organising information. They provide unique identifiers - in the form of URIs - allowing to connect data coming from different digital sources, archives, and libraries more easily. Gazetteers are also a subject of research in their own right, helping us to define and analyse our geographical knowledge of the past, and interrogate the nature of what we call “a place”[25].

The following types of gazetteer datasets were considered in assessing use cases:

- Geographical names
- Addresses
- Administrative units
- Buildings

[25] https://pelagios.org/case-studies/what-are-urban-gazetteers/

In general, 'Geographical Names' data is used mainly for identifying and locating geographical features, i.e. as search criteria in gazetteer services, geo-portals, geo-catalogues, etc. In other words, one of its main uses is for the geocoding process[26]. Geographical names are also a key element of any kind of map: no one would understand a map without geographical names. This applies for any kind of graphical representation (background 2D maps, ortho-images, 3D models). This theme is of great interest for many Sustainable Development Goals (SDGs) related use cases: in the analysis phase or for more operational purposes, such as locating where the people are (e.g. to ensure them accessibility to services or to assess the human pressure on environment or to assess the impact of a risk or pollution), the data on named places is useful.

Box 1. Geographical names - general use cases	Emergency response Economic, social and environmental analysis Cultural identity and heritage Mapping and navigation Providing a link / index function to other spatial and non-spatial data

Addresses provide one of the most common ways to determine a physical object for purposes of identification and location; assisting services such as postal delivery, emergency response, marketing, mapping, utility planning and land administration. The most obvious use of addresses is the physical delivery of mail, but the power of address data in the digital age lies in its geocoding capability[27]. A great deal of information is linked to addresses, and a geocoded address database allows such information to be linked to a physical location. Thus, address data can add significant economic value, for example in marketing and logistics. Address data underpins government administration at all levels, and good administration is a prerequisite for the achievement of the SDGs. It supports the provision of services and also enables effective communication with citizens: informing them of policies applying to them, notifying them of events or incidents affecting them and supporting the carrying out of social surveys.

Box 2. Addresses - general use cases	Geocoding of statistical surveys, manage emergency rescue, locate where people are, accessibility studies, manage incidents; locate economic activities in ecosystem accounting

Administrative units determine unambiguously the responsibilities and competences of the various authoritative entities in relation to any area of a Member State[28]. In the analysis phase, any government has to know the geographic extent for its expected actions. Administrative units are widely used in the management of geographic information, for instance to "crop" other datasets as delivery units, and are often based on the country administrative division or as search criteria in gazetteer services, GeoPortals, GeoCatalogues etc. At national level, municipalities are generally used to build the cadastral system and administrative unit names are also the basis for the address system.

[26] https://un-ggim-europe.org/wp-content/uploads/2018/11/UN-GGIM-Europe_WGA_Recommandation_Content-GN-v1.0.pdf
[27] https://un-ggim-europe.org/wp-content/uploads/2018/11/UN-GGIM-Europe_WGA_Recommandation_Content_AD-v1.0_0.pdf
[28] https://un-ggim-europe.org/wp-content/uploads/2018/11/Recommandation_Content_AU-v1.0.pdf

Box 3. Administrative units - general use cases	Mapping or use as statistical units, manage emergency rescue, waste management plans, protect water ecosystems, find responsible party for policy implementation and administration, forest management, subsidies for farmers, forecast agricultural production, spatial planning, monitoring of regional and urban policy implementation using territorial typologies based on administrative units, maritime spatial planning, integrated coastal management

Buildings are 3D topographic objects and, as such, may influence the propagation of physical phenomena[29]. Building data is also sometimes needed to make visibility or inter-visibility maps, to forecast how noise will propagate, etc. At the same time, the physical phenomena may also impact buildings. The most obvious use case is the assessment of the vulnerability of buildings to various kinds of risks (earthquake, fire, flood, etc.), according to the physical characteristics of the building. It is also of significant interest to assess the ability of the buildings to contribute to SDG related improvements, e.g. can the roof host solar panels? Can the building be isolated from noise or from heat losses?. In addition, buildings are valuable economic assets and for some of them, part of historic patrimony; that should be taken into account in risk management.

Box 4. Buildings - general use cases	These data are required for serving citizens (e.g. school, hospital), assessments for air and noise pollution or risk assessments to various kinds of risks (earthquake, fire, flood, etc.), monitoring of land consumption, population concentration and access to services, and are crucial for the emergencies

There are also some other important gazetteer datasets used in various services provided by the public and private sectors. However, they are not in the scope of this report.

Part of the work was dedicated to identifying the Europe-wide use cases where there is a demand for different datasets or services involving gazetteer data. Of particular interest, were cross-border (between neighbouring countries) or transnational (between different countries) use cases, such as:

- World Historical Gazetteer
- Le Grande Région/Die Großregion
- Centrope Region
- The Locator – Business information for enterprises
- Energinet
- Applying for a fiscal number scenario

With the exception of the first and the last case, interviews were performed with representatives of the use cases discussing with them different aspects of (potential) use of gazetteer datasets and services:

- Type of datasets/services
- Sources of datasets/services (authoritative, open, voluntary, API, …)
- Compliance to standards (e.g. INSPIRE)
- Experience and obstacles encountered in cross-border data/service acquisition
- Licencing policy

[29] https://un-ggim-europe.org/wp-content/uploads/2018/11/UN-GGIM-Europe_WGA_Recommandation_Content_theme-BU-v1.0.pdf

- The need for common pan-European gazetteer services
- The need for historical datasets

The goal of interviews was to cover as much as possible different corners of Europe, and on the other side to involve stakeholders from different sectors.

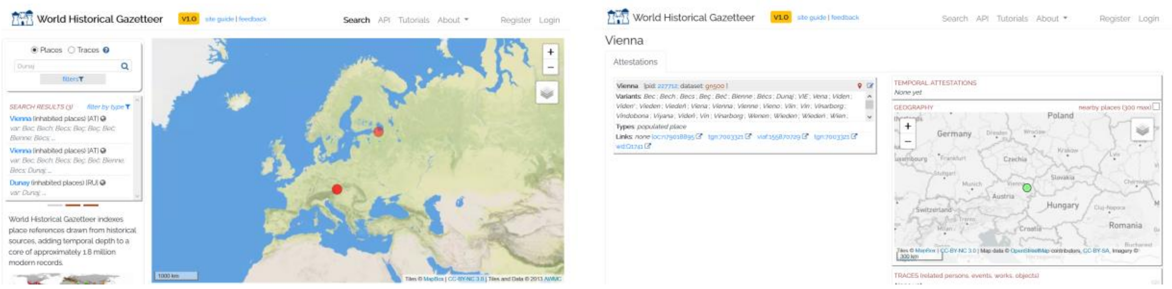
Sections 2.3.1 to 2.3.6 are covering short introduction of all cases, while section 2.3.7 presents the summary of the performed interviews of cases 2.3.2 to 2.3.5.

2.3.1 World Historical Gazetteer

The World Historical Gazetteer (WHG), because of its global coverage, is out of the scope of this report. However, since it is an excellent example of a global cooperation in building a gazetteer, respecting the historical and multilingual aspects, it is included and discussions with its creators have taken place.

The WHG (**Figure 26**) provides a collection of content and services that permit world historians, their students, and the general public to do spatial and temporal reasoning and visualisation in a data rich environment at global and trans-regional scales. The WHG index has been seeded with core data from several essentially modern sources, to which first historical contribution has been added, providing initial broad coverage of modern places, adding then historical depth over time via contributions of attestations from historical sources. WHG belongs to a growing community of project teams and individuals working to link data about the past, principally by linking data about historical places, and encouraging communities of interest for particular regions, periods, and themes to establish specialised "domains". This approach has worked effectively for scholars of the Ancient Mediterranean and the Pleiades[30] and Pelagios[31] projects.

Figure 26. World Historical Gazetteer



Source: WHG[32]

2.3.2 Le Grande Région/Die Großregion

Le Grande Région/Die Großregion[33] lies at the crossroads of the rivers Rhine, Saar, Meuse and Moselle. It covers 65.401 km2 with more than 11.6 million inhabitants from the territories Lorraine in the French region Grand Est, Wallonia, the Federation Wallonia-Brussels and Ostbelgien in Belgium, Saarland and Rhineland-Palatinate in Germany as well as the Grand Duchy of Luxembourg.

The concept of the Greater Region has its origins in the intergovernmental commission set up by Germany and France in 1969. Two years later, the Grand Duchy of Luxembourg joined the commission, followed by the German federal states Saarland and Rhineland-Palatinate, the French region Lorraine as well as the Federation Wallonia-Brussels and the German-speaking Community of Belgium. Over the years the member countries established an institutional framework for cross-border cooperation. In 1995, the first Summit of the Executives of the Greater Region took place in Mondorf-les-Bains (LU). Since then, the cooperation partners have intensified their efforts to tackle common challenges in the areas of:

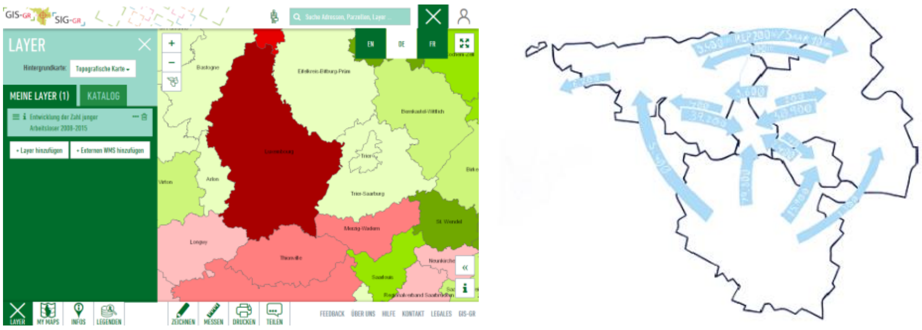
- Mobility and regional development
- Education and lifelong learning

[30] https://pleiades.stoa.org/
[31] https://pelagios.org/
[32] http://whgazetteer.org/
[33] http://www.granderegion.net/

- Economy and competitiveness
- Society and security
- Tourism and culture
- Environment and sustainability

The Greater Region offers attractive and multilingual working and living conditions for more than 11.6 million inhabitants and approximately 250,000 cross-border commuters per day – the highest number of cross-border commuters in Europe.

Figure 27. Le Grande Région/Die Großregion



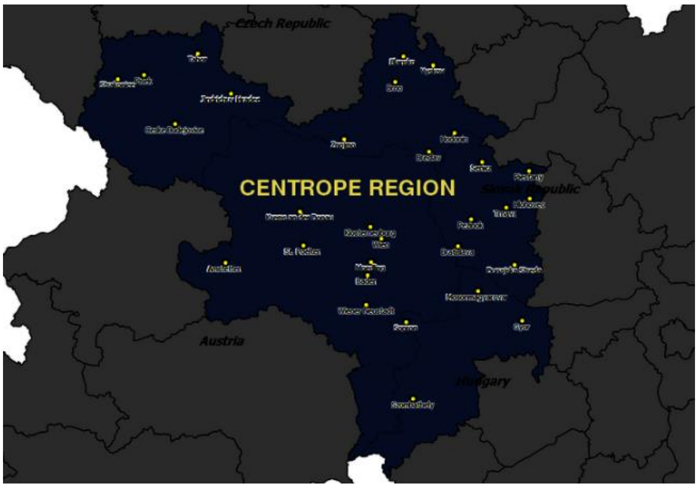
Source: Le Grande Région/Die Großregion[32]

A geoportal has been set up for the Greater Region which enables users to view the majority of the maps created by GIS-GR in the form of cross-border layers on an interactive map. The map application is developed in collaboration with the Luxembourg Cadastral Office.

2.3.3 Centrope Region

The Centrope Region (**Figure 28)** consists of a number of adjacent counties and states along the borders between Austria, Czechia, Hungary, and the Slovak Republic, comprising regions Jihomoravský, Bratislavský, Trnavský, Győr-Moson-Sopron, Burgenland, Lower Austria, and Vienna. It was founded in 2003 by the political declaration of Kittsee and aims at strengthening partnership and economy in a region which was affected by the Iron Curtain during the second half of the 20th century.

Figure 28. Centrope Region



Source: Centrope Region[34]

[34] www.centropemap.org/

For many years, the people of the region have intermingled. They make daytrips or work in the neighbouring countries. As a "figurehead" for regional cooperation, Centrope mirrors the commonality which already has become part of everyday life and a matter of normalcy for its inhabitants.

Within the framework of the Interreg IIIA project "Building a European Region", the period until the beginning of 2006 served to create the necessary groundwork – cooperative structure and services – to enable the establishment of the CENTROPE Central European Region.

CentropeMAP[35] is a geoportal connecting the region by collecting web map services from the partner countries. The services are brought together in a single map viewer and allow the user to experience a cross-border working area with dozens of data layers dealing with all topics which could be of interest for regional planners and similar professionals. Layers are coming from the fields of biota, boundaries, elevation, imagery/base maps, inland waters, planning/cadastre, structure, and transportation.

2.3.4 The Locator – Business information for enterprises

'The Locator' (**Figure 29**) evolved from the INTERREG IVA – project 'Industrial Site Portal EMR' with the financial assistance of the following public bodies: The State of North Rhine-Westphalia (DE), Province of Limburg (NL), Province of Limburg (BE), Walloon region (BE).

'The Locator' is cross-border system for providing information on business locations. Companies setting up in business in the Euregion Meuse-Rhine (Germany-Netherlands-Belgium) for the first time or wishing to relocate or expand can use it to find up-to-date business location information and make cross-border comparisons.

Information on commercial sites and property is not only available online but can also be downloaded in the form of clear location reports. In addition, users receive a detailed summary of businesses and sectors already based in the region. Also provided is basic information which is essential for companies setting up a business in the region, e.g. information on the labour market and taxes or contact details of individual partners.

Figure 29. The Locator

Source: The Locator[36]

"The Locator" is a four-language, multifunctional information system for enterprise locations in the Euregio Meuse-Rhine. It is the first of its kind in Europe and creates transparency across national borders. It doesn't matter if you are looking for information about industrial space, business properties, enterprises or useful tips for your relocation: you will find everything you need to know about the future location of your enterprise.

"The Locator" uses SOAP interfaces to the different regional data sources and thus follows the so-called "data-at-the-source" principle. Therefore, "The Locator" is up to date. For data on "business parks" and "real estates" synchronisation / update happens once a day. "Company" and "Settlement" data are adapted as required when substantial changes occur.

[35] http://map.centropemap.org/
[36] http://www.the-locator.eu/

2.3.5 Energinet

Energinet[37] is an independent public enterprise owned by the Danish Ministry of Climate, Energy and Utilities, owning, operating and developing the transmission systems for electricity and natural gas in Denmark.

Energinet's European activities are carried out in collaboration with the Danish Ministry of Energy, Utilities and Climate and through the two umbrella organisations for the electricity and gas sectors: ENTSO-E and ENTSOG.

European Transmission System Operators (TSOs) have collaborated for decades, a collaboration that most recently was formalised in the European Commission's third liberalisation package which also established the two ENTSOs in Brussels. Energinet is now an active participant in both ENTSOs.

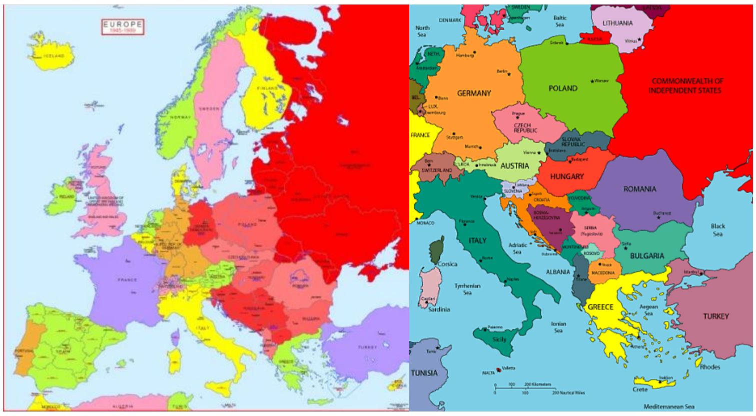
Energinet has a long tradition for regional cooperation in the electricity and gas sector. Energinet continues to develop regional cooperation which works as a steppingstone for European collaboration. The Nordic countries are in the process of establishing a joint office, where a group of specialists from the Nordic TSOs will work to expand collaboration between the Nordic countries to reach solutions for a number of operational tasks.

Energinet has a number of energy datasets, which are available for other parties. Most data are open for everybody, other data requires login. One of the features is an interactive map showing real time Energy production, consumption, import and export to and from DK. Additionally, maps are available for both Electricity and Gas which include historical time series.

2.3.6 Applying for a fiscal number

This scenario emerged from experiences with some of JRC staff when arranging the formalities with authorities related to moving from the country of residence to Ispra, IT, for the period of a working contract. Amongst others, each Italian non-resident (applicant) has to obtain a unique identifier called a "codice fiscale" (fiscal number identification), which serves as an identification in many life events in communication with public authorities as well in business or other commercial processes. That unique identifier is calculated based on an algorithm with input of many parameters. One of the input parameters is the country of birth of the applicant. Since the 1990s, there have been many changes in terms of country names in Europe; some of them dissolved (Czechoslovakia, Yugoslavia, Soviet Union, Federative Republic of Yugoslavia, Serbia and Montenegro), some of them integrated with other countries (German Democratic Republic), and some of them have changed their names (FYR Macedonia).

Figure 30. Changes in Europe



Source: https://www.euratlas.net

In these situations, the public authority providing the identifier needs to have historical lineage of the data set providing the administrative borders and names of the countries. This is not always possible, since some authorities do not have the historical lineage of the datasets available but rather only their latest status. That could lead to the situation that different authorities could provide different unique identifiers for the same person (applicant).

[37] https://en.energinet.dk/

2.3.7 Summary of the use cases

Looking at the use cases described in sections 2.3.2-2.3.5 and the feedback from use case representatives on the aspects of (potential) use of gazetteer datasets and services listed in section 2.3, the following conclusions can be drawn:

- In all of the cases, some gazetteer datasets or services are used or are planned to be used, to support the basic processes.
- There is no common way of using available data sources and services. The approach depends on the availability of data in the regions (NUTS2, NUTS3, …), in fact many different possibilities exist and thus datasets from various sources are used in the examples above – authoritative, open voluntary, commercial sources of data and services and even own data acquisition.
- One of the major challenges in the process of data/services acquisition is the variety of different sources from different countries, which leads to a greater need for data harmonisation and interoperability in integrating data for use in common systems.
- Another issue, having in mind different available data sources in different countries, is the licensing conditions under which data and services can be reused, which can also represent a great challenge to the builders of cross-border systems.
- Few large-scale datasets are compliant with the INSPIRE data specifications. However, the situation is better in relation to services, where WMS and WFS compliant services are used and in the future the use of the OGC API standards is expected.
- In all cases, the multilingual and historical aspects of datasets are considered, though the importance varies from case to case.
- As a general conclusion, based on these use cases, a common EU gazetteer data service(s) would be of value in relation to:
    - Detailed and accurate information on addresses
    - Historical information on administrative units
    - Buildings and their function of use
    - Points of Interest for tourism purposes (multilingual).

3 Conclusions and next steps

In recent years, use cases and consequently demand for spatial data have evolved. The number of spatial analyses has increased and this is now common practice in many services of the European Commission. Spatial analyses support policy making in domains such as environment, statistics, agriculture and transport, or other community data programmes such as Copernicus. In addition, the adoption of the UN 2030 agenda for Sustainable Development[38] and of the related indicator framework for monitoring the Sustainable Development Goals (SDGs) also result in greater demand for various types of data, including spatial data. In response to these growing needs, several Directorates-General (DG) of the European Commission and European Agencies have therefore agreed that the Commission needs more and better-quality geospatial data from official sources with European or at least EU coverage. The focus is on official data that Member States normally should provide to the Commission under several European legal obligations, including the INSPIRE, Public Sector Information and Intelligent Transport Systems Directives. However, in the absence of official data, volunteered geographic information such as OSM is considered as well as a valuable source of spatial data.

In the context of digital public services and their underlying business processes, a lot of data and information are consumed, transmitted and processed in order to create new information. In most of these processes there is often a mixture of different types of data, including location information. Authentic or base registers will often play a crucial role, e.g. addresses, cadastral parcels, buildings, but also persons and businesses.

The INSPIRE Directive is a major effort to harmonise and share data between administrations, and also cross border. This will enable the better sharing of environmental spatial information among public sector organisations and better facilitate public access to spatial information across Europe, which will benefit citizens and businesses alike. As the EU SDI is built on the national SDIs, the aspect of full re-use of national SDIs is very strongly embedded in the Directive.

In this report two datasets, official and volunteered, were analysed on a pan-European scale: Geographical Names and Addresses, both of which are part of INSPIRE Annex I data themes and are considered as reference data. These datasets are also recognised by UN-GGIM: Europe and are part of UN-GGIM: Europe Core Data themes[39]. The results of the analysis showed that both types of datasets, official and volunteered, are complementary and mutually enhancing results can be obtained by combining the two sources.

Geographical names are used extensively when searching for information in web-services (including geoportals), navigating, referencing thematic information to a location (geocoding), visualising geographic information on maps and screens, as well as when processing spatial datasets comprising historical data. Correct usage of geographical names is a principal aspect of everyday communication; consequently, the status (official, historical, etc.) and linguistic properties (language, spelling, eventual transliteration, etc.) are a prime interest of many users, including press agencies, map publishers, spatial analysts, authorities, etc.

Address data underpin government administration at all levels. Addresses support the provision of services and also enable effective communication with citizens: informing them of policies applying to them, notifying them of events or incidents affecting them and supporting the carrying out of social surveys.

The study has developed its own "Cultural Heritage Testbed" and examined a series of existing cross-border and pan-European use cases to determine the possible requirements for a pan-European gazetteer. The conclusion of these discussions is that a common EU gazetteer data service(s) is of value and should enable (at least):

- Detailed and accurate information on addresses
- Historical information on administrative units
- Buildings and their function of use
- Points of Interest for tourism purposes (multilingual).

To further cement this view, the upcoming implementing act on High Value Datasets, defined under the scope of the Open Data Directive (ODD), will define a series of open datasets in various categories (including geospatial data). At the time of writing this report, consultations are still taking place. Nevertheless, four categories of 'gazetteer' datasets, namely addresses, administrative units, geographical names and buildings, are being considered. This will 'open up' authoritative gazetteer data in Europe but will not necessarily require harmonisation of datasets. Harmonisation of data themes is however covered under the INSPIRE Directive. The combination of the two Directives, therefore, changes the possibilities in relation to future pan-European gazetteer services.

[38] https://www.un.org/sustainabledevelopment/development-agenda/
[39] https://un-ggim-europe.org/working-groups/working-group-core-data/

A further consideration will be the European Data Strategy, which envisages the implementation of a series of common thematic European data spaces using a federated cloud infrastructure. Location data will be important for all the data spaces and a harmonised approach on reference data such as the gazetteer datasets will generate efficiencies and facilitate better outcomes in line with the aims of the strategy.

Whether the demand for European gazetteer services is met by the public sector, the private sector or other community services is yet to be seen. Certainly, the ODD changes the potential relationship between European public administrations and OpenStreetMap. Both ODD and the European data strategy also aim to facilitate better links between the public and private sector.

Moreover, the proposed data categories in the foreseen implementing act on High Value Datasets, defined under the Open Data Directive, have a clear overlap with the data scope of the INSPIRE Directive and could benefit from reusing already available spatial data in the INSPIRE infrastructure.

The clear future goal is the development of EU gazetteer services covering EU Member States and beyond, based primarily on authoritative data, but also complemented with open volunteered data.

References

Cetl, V.; Tomas, R.; Kotsev, A.; de Lima, V.N.; Smith, R.S.; Jobst, M. Establishing Common Ground Through INSPIRE: The Legally-Driven European Spatial Data Infrastructure. In Service-Oriented Mapping; Springer, 2019; pp. 63–84.

EN ISO 19112:2019, Geographic information - Spatial referencing by geographic identifiers (ISO 19112:2019), Comite Europeen de Normalisation, 2019.

European Parliament and European Council. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal of the European Union 2007, L 108, 1–14.

Pignatelli, F., Boguslawski, R., Fernández de Soria, A., Gielis, I., Bargiotti, L., Goedertier, S., *Feasibility study for an EU Gazetteer common service*, JRC Technical Report, Ispra, European Commission, 2017.

UNGEGN 2007, Technical reference manual for the standardization of geographical names, United Nations Group of Experts on Geographical Names, 2007, ISBN: 92-1-161500-5.

List of abbreviations

AD	Addresses
DG	Directorate-General
ELISE	European Location Interoperability Solutions for e-Government
ESTAT	Eurostat
EU	European Union
EC	European Commission
GISCO	Geographic Information System of the Commission
GN	Geographical Names
INSPIRE	Infrastructure for Spatial Information in the European Community
ISA	Interoperability solutions for public administrations
ISA^2	Interoperability solutions for public administrations, businesses and citizens
ISO	International Organization for Standardization
SDGs	Sustainable Development Goals
SDI	Spatial Data Infrastructure
JRC	Joint Research Centre
MS	Member States
ODD	Open Data Directive
OSM	OpenStreetMap
SDGs	Sustainable Development Goals
UN	United Nations
UNGEGN	United Nations Group of Experts on Geographical Names
UNGGIM	United Nations Global Geospatial Information Management

List of boxes

List of figures

List of tables

KJ-NA-30378-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office
of the European Union

doi:10.2760/095025

ISBN 978-92-76-22301-6