EuroSDR

European Spatial Data Research

February 2023

Geodata Discoverabiltity

Joint Workshop of EuroGeographics and EuroSDR

April 28th 2022, Online

Bénédicte Bucher, Marcin Grudzień, Nathalie Delattre, Jordi Escriu Paradell, Erwin Folmer, Antonin Garrone, Antje Kügeler, Ángel Lopez, Ed Parsons, Andrea Perego, Jiri Pilar, Jari Reini, Hannes I. Reuter, Jill Saligoe-Simmel, Maja Schneider, Jeroen Ticheler

Official Workshop Report

Published by EuroSDR

EUROPEAN SPATIAL DATA RESEARCH

PRESIDENT 2020 – 2023:

Michael Hovenbitzer, Germany

VICE-PRESIDENT 2021 – 2023:

Fabio Remondino, Italy

SECRETARY – GENERAL 2019 – 2023:

Joep Crompvoets, Belgium

DELEGATES BY MEMBER COUNTRY:

Austria: Wolfgang Gold, Gottfried Mandlburger
Belgium: Eric Bayers
Croatia: Željko Bačič
Cyprus: Elikkos Elia, Georgia Papathoma, Andreas Hadjiraftis, Dimitrios Skarlatos
Denmark: Jesper Weng Haar, Tessa Anderson
Estonia: Tambet Tiits, Artu Ellmann, Evelyn Uuemaa
Finland: Juha Hyyppä, Juha Kareinen
France: Bénédicte Bucher, Aurélien Plyer
Germany: Michael Hovenbitzer, Norbert Haala
Ireland: Paul Kane, Audrey Martin
Norway: Jon Arne Trollvik, Ivar Maalen-Johansen
Poland: Anna Bober, Krzysztof Bakuła
Portugal: Mário Caetano, Rui Reis, Marisa Silva
Slovenia: Dalibor Radovan, Peter Prešeren, Marjan Čeh
Spain: Julián Delgado Hernández
Sweden: Tobias Lindholm, Anders Rydén, Heather Reese
Switzerland: André Streilein
The Netherlands: Martijn Rijsdijk
United Kingdom: Claire Ellul

ASSOCIATE MEMBERS AND THEIR REPRESENTATIVES:

Digitaal Vlaanderen: Jo Van Valckenborgh
Esri: Nick Land, Konrad Wenzel (nFrames)
Field Group: Leif Erik Blankenberg
Hexagon: Simon Musaeus
ICGC: Julià Talaya
IGI: Philipp Grimm
RIEGL: Peter Rieger
Vexcel: Michael Gruber
1Spatial: Dan Warner

COMMISSION CHAIRPERSONS:

Data Acquisition: Jon Mills, United Kingdom
Modelling, Integration and Processing: Norbert Haala, Germany
Information Usage and Visualization: Martijn Rijsdijk, The Netherlands
Business Models and Operations: Frédéric Cantat, France
Knowledge Transfer: Anka Lisec, Slovenia

OFFICE OF PUBLICATIONS:

Bundesamt für Eich- und Vermessungswesen
Publications Officer: Christine Ressl
Austria
Contact via EuroSDR secretariat

CONTACT DETAILS:

EuroSDR Secretariat
KU Leuven Public Governance Institute
Faculty of Social Sciences
Parkstraat 45 bus 3609
3000 Leuven
Belgium

Tel.: +32 16 37 98 10
Email: eurosdr@kuleuven.be
Web: www.eurosdr.net

The official publications of EuroSDR are peer-reviewed.

Bénédicte Bucher, Marcin Grudzień, Nathalie Delattre, Jordi Escriu Paradell, Erwin Folmer, Antonin Garrone, Antje Kügeler, Ángel Lopez, Ed Parsons, Andrea Perego, Jiri Pilar, Jari Reini, Hannes I. Reuter, Jill Saligoe-Simmel, Maja Schneider, Jeroen Ticheler

"Geodata Discoverability"

Joint Workshop of EuroGeographics and EuroSDR, 28th April 2022 - Online

Table of contents

GEODATA DISCOVERABILITY

Joint Workshop of EuroGeographics and EuroSDR

April 28th 2022 - Online

With 23 figures and 1 table

Bénédicte Bucher [a], Marcin Grudzień [b],
Nathalie Delattre [c], Jordi Escriu Paradell [d], Erwin Folmer [e,f], Antonin Garrone [g],
Antje Kügeler [h], Ángel Lopez [i], Ed Parsons [j], Andrea Perego [k], Jiri Pilar [l],
Jari Reini [m], Hannes I. Reuter [n], Jill Saligoe-Simmel [o],
Maja Schneider [p], Jeroen Ticheler [q]

[a] National Institute for Geographical and Forest Information (IGN)
Saint Mandé, France
Benedicte.Bucher@ign.fr

[b] Head Office of Geodesy and Cartography (GUGiK)
Warsaw, Poland
Marcin.Grudzien@gugik.gov.pl

[c] National Geographic Institute of Belgium
Brussels, Belgium

[d] European Commission Joint Research Centre (JRC)
Ispra, Italy

[e] University of Twente,
Enschede, the Netherlands

[f] Kadaster
Apeldoorn, the Netherlands

[g] Etalab France
Paris, France

[h] Con Terra
Münster, Germany

[i] European Center for Medium-Range Weather Forecasts (ECMWF)
Reading, UK

[j] Google
London, UK

[k] European Parliament
Luxembourg, Luxembourg

[l] European Commission
Luxembourg, Luxembourg

[m] National Land Survey (NLS) of Finland,
Helsinki, Finland

[n] Eurostat
Luxembourg, Luxembourg

[o] Esri
Indianapolis, US

[p] Technical University of Munich
Munich, Germany

[q] GeoCat
Ede, the Netherlands

Index of Figures

Index of Tables

1 Introduction

Geodata discoverability was identified decades ago as a crucial challenge both by the Spatial Data Infrastructure (SDI) community and, for the more general case of data discoverability, by the Web of data community. It has kept growing in the late years.

SDIs adopt distributed infrastructures to lower the cost of data acquisition by reusing available data instead of producing new ones (Masser 1999). A core function on distributed infrastructures is the discovery of available data, initially addressed through catalogue services. Another motivation for SDIs is to ensure trust by using data from the relevant authorities like, for example, reusing member state data to establish and monitor European policies. This also implies that member state data that are relevant to a European decision can be discovered ahead of this decision. The INSPIRE SDI was established by the European Commission for supporting the sharing and reuse of member state data produced by legally mandated authorities and relevant to European environmental policies (EP 2020).

The Web of data is more focused on increasing the value of data by allowing new usages beyond their initial purpose. Hence, data should be visible from developers with diverse backgrounds and levels of expertise in the domain of these data (Berners Lee et al. 2001).

This stake of data discoverability has kept growing lately for many reasons.

First reason is the growing amount of available data. With the development of sensors, of social networks and the digitisation of processes, our societies produce more and more data. Let us mention some examples, such as data acquired by satellites, or data produced by collaborative online communities. Data is also more published, following the development of open data policies to improve the transparency of administration and, more generally, as a paradigm to support trust and innovation. In Europe in particular, the amended version of the Public Sector Information Directive is entitled Open Data Directive (Directive (EU) 2019/1024).

Second reason is the rising awareness that data is an opportunity to better cope with the complexity of our world during decision-making. For example, at the international level, a set of sustainable development indicators has been proposed by the United Nations to cover all possible aspects of UN Agenda 2030 like, for example, access to water, to food, to education, prevent war, no poverty, and inclusive cities. The main motivation for these indicators is to monitor quantitatively how our planet evolves. Another purpose that they also serve is to gather yearly data to understand the interdependencies between these different dimensions of sustainability. All in all, in order to describe our world complexity that is relevant to a decision at stake, one may need data with different thematic focuses and different levels of detail, possibly also different spatial and temporal scopes. These data usually refer to different platforms, providers, technologies, and communities, which makes their discovery complex from the user perspective.

With these considerations above, it is no surprise that actors beyond SDI and Web of data specialists, from the public or private sector, with different backgrounds and ambitions at local or global level, face dataset discovery tasks or engage in the design of discovery solutions.

In April 2022, EuroSDR and EuroGeographics jointly organised a workshop to share practical experiences, findings and roadmaps related to the discoverability and accessibility of data with a geographical characteristic, including data covered by the INSPIRE directive. EuroSDR is a not-for-profit organisation, established in 1953, linking National Mapping and Cadastral Agencies with Research Institutes, Universities and Companies in Europe to mutualise innovation efforts in spatial data provision, management and delivery. Eurogeographics is an association that gathers National Mapping and Cadastral Agencies, to share and resolve common problems, to manage pan-European projects such as web-based metadata services and integrated geographic databases, and to become the unified “voice of European National Mapping, Cadastral and Land Registry Authorities”. This publication reports the main findings from this workshop.

2 Creation of pan-European geospatial datasets – Experiences in real-life

Hannes Reuter, Eurostat

In the Directorate of Eurostat, a permanent service (GISCO) of the European Commission is installed that answers the needs of Eurostat and the European Commission for geographical information at the level of the European Union (EU), its Member States and regions. To bring Europe closer to citizens and regions, Eurostat continues to make use of georeferenced data by systematically integrating and mainstreaming geospatial information into statistical production and collecting more geospatial data, with a special focus on the core geospatial datasets, for detailed regional and local level statistics. Therefore, one part of GISCO's activities is the creation of pan-European datasets based on authoritative data providers to support activities outlined in the Global Statistical Geospatial Framework, illustrated figure 1.

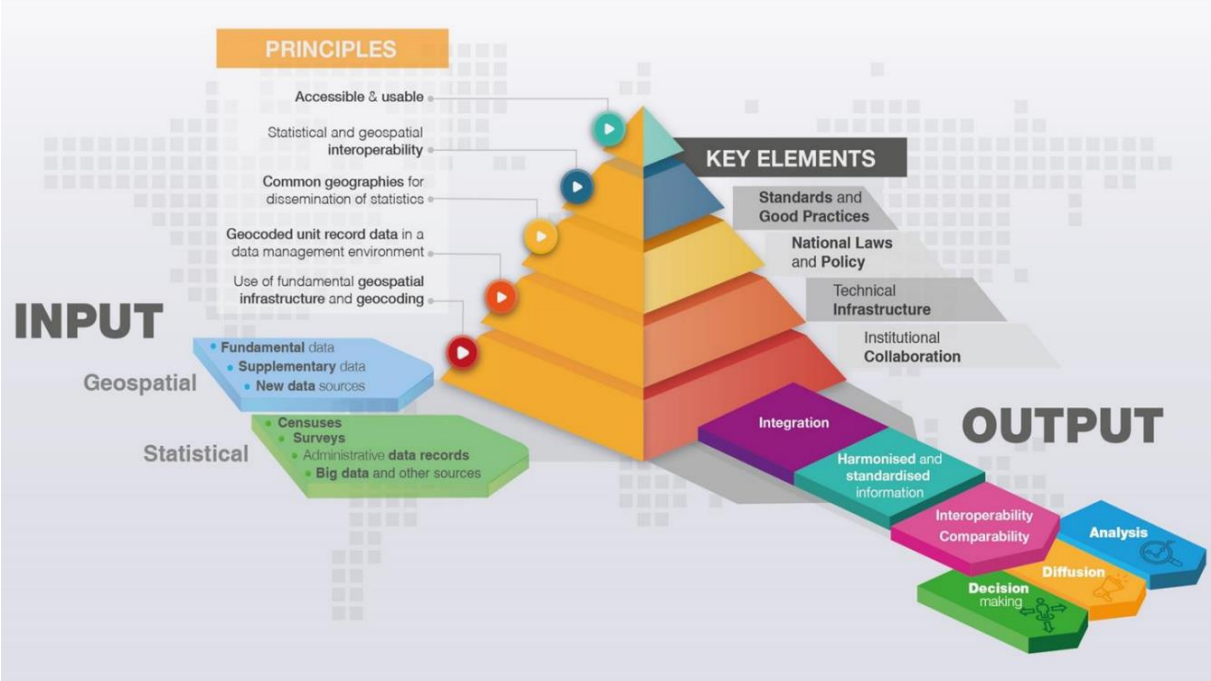


Figure 1: Global Statistical Geospatial Framework, © UN-GGIM

Datasets are identified in various computer systems (e.g. data.europa.eu). Unfortunately, in some countries and for some themes, the GISCO team fails to discover any data through these global catalogues, portals and retrieval solutions. Human intelligence (e.g. via local contacts, social media) is used to complete the information for a given data theme and country. These datasets are then analysed, ingested, transformed, quality assessed and quality controlled. Possible feedback rounds with data provider organisations are initiated with the final aim of a pan-European dataset, see figure 2. Services are built on top to support policy analysts in the European Commission or wider public organisations.

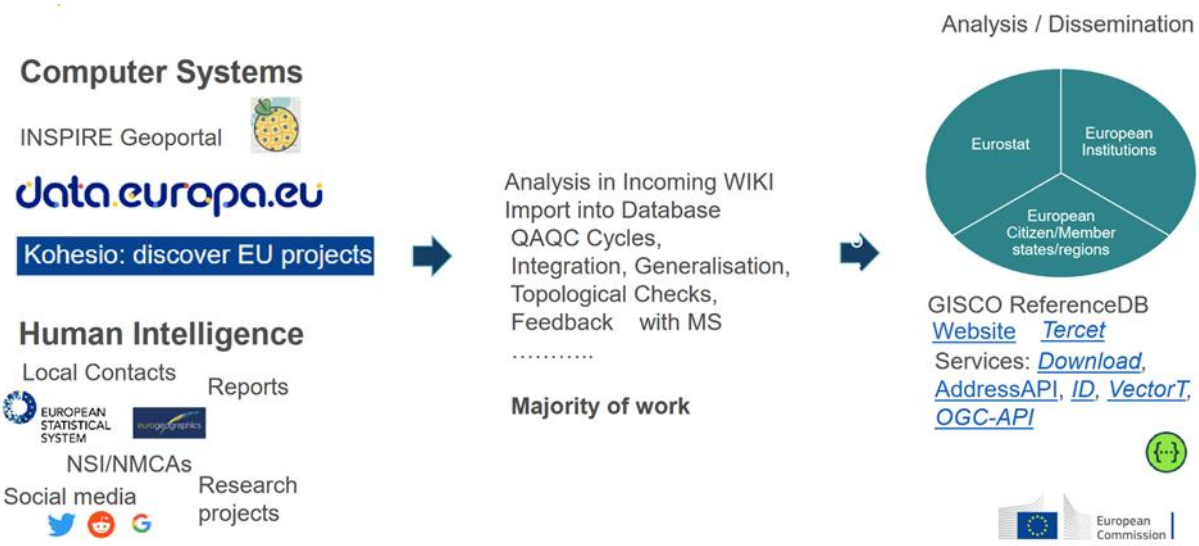


Figure 2: Workflows for establishing pan European datasets, ® Reuter, Eurostat

GISCO has observed some of the following issues while creating pan-European dataset: a) data exists but are not documented in metadata systems (neither European or Member States) or directly hidden, b) several datasets with various specifications exits for the same theme, however are maintained by different organisations, c) dispersed datasets per country without a central authority exists, d) depending on organisations administrative hurdles or a lock down mentality without user centric views are shown, e) officially data are declared to not exists, but commercial datasets exists for the same area, and f) outdated (temporally) and semantically reduced data (e.g. content) are provided officially while rich and timely data exists.

Eurostat is on a good way in creating pan-European datasets for policy purposes and operational results are promising for a series of themes. The progress is hampered by specific countries due to a) non-documented and available datasets, b) stability of services, c) performance of services or d) use restrictions due to financial demands/legal provision and e) quality of geospatial data provided.

3 Discoverability of Open Maps for Europe data

Nathalie Delattre, IGN-Belgium

Open Maps for Europe (OME), https://www.mapsforeurope.org/, is a project co-funded by the European Commission, to deliver free to use maps from more than 40 European countries, as illustrated in figure 3.

The purpose is to provide easy access to free-to-use harmonised maps for Europe and realise the benefits of the Open Public Sector Information directive. Eurogeographics has developed expertise in producing pan-European datasets using official maps, geospatial data and land information from authoritative national sources. An important achievement of the project is to confer an open licence to these pan-European datasets, to make the maps discoverable, viewable and downloadable through an enhanced gateway.

Figure 3: Products that compose Open Maps for Europe, © Delattre, NGI

Regarding lessons learnt during the project OME, dealing with different metadata encoding and format (INSPIRE and OPEN data) might provide some burden because this duplicates the effort for the provision of the same information. A strategy needs to be established for efficient workflows of metadata harvesting between the different portals to lower the workload. This strategy should favour consistent metadata encoding among the portals (INSPIRE and open data, national and European). One solution is that metadata for geospatial information follows INSPIRE standards and that open data platforms adopt these or be able to harvest them without information loss. Valuable conversion tools should therefore be implemented.

4 The experience of the French open data portal

Antonin Garrone, Etalab France

Etalab is a French administration created in 2011 to foster the development of data based public action. Its main actions are organised around three axes : 1) opening data (data.gouv.fr, api.gouv.fr), 2) exploiting data through data science and artificial intelligence 3) opening public action through “public interest entrepreneurs”. “Public interest entrepreneurs” are short term missions taken by civil servants and by salaries from the private sectors.

Data.gouv.fr is the single point of entry for national open data in France, accessing the data themselves or referencing portals from specific providers. It has no specialisation in geographical data unlike other eco-systems but rather in open data published by the public sector. With respect to geographical data, a former specific platform was dedicated to geographical open data, geo.data.gouv.fr, which harvested more than 66 catalogues. It is currently finished and shall merge within the INSPIRE French Geocatalogue operated by BRGM.

Everyone is welcome to publish on the portal both datasets and reuse. Moderation is done post-publication by Etalab.

In that context, the multiplicity of models and issues in fields correspondences of these models (DCAT, ODS, CKAN, UDATA, etc.) can jeopardise data discoverability, as well as possible low quality of metadata.

A key topic is thus the data models interoperability. Schema.data.gouv.fr has been designed to create and reference data schemas in order to enhance discoverability, reusability and interoperability of data on data.gouv.fr. This key aspect is illustrated on figure 4. It is associated with publier.etalab.studio, a

tool to edit, validate and publish data structured according to a schema. Examples of schemas are : the National Address Database and the Infrastructure for electric vehicle recharge.

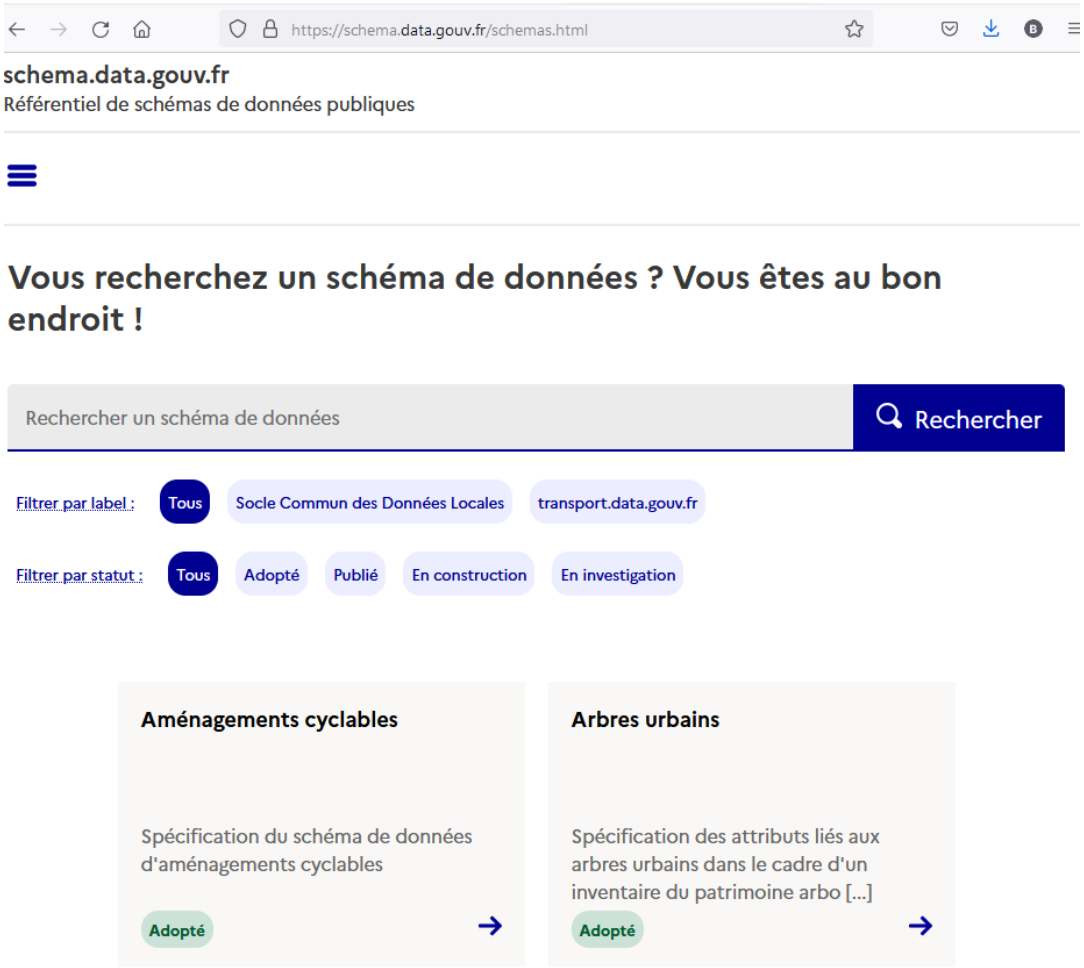


Figure 4: Snapshot from schema.data.gouv.fr, a portal to support the creation, publication or reuse of data models.

Last, to enhance data discoverability, etalab prepares and curates specific outputs to present directly to users ‘essential’ geographical datasets.

5 Discoverability of open agriculture data

Maja Schneider, Technical University of Munich

Within the EuroCrops project, open agricultural data from member states of the European Union has been collected over the past one and a half years. Most notably, farmers’ self-declarations, including the georeferenced polygon with the crop information gathered for the subsidy control, have been the primary focus for the project. In last years, an increasing number of ministries and paying agencies have decided to make this data publicly available.

However, the theoretical availability of the data does not necessarily make it straightforwardly accessible because of heterogeneities of data release practices, as illustrated in figure 5. While platforms such as the INSPIRE Geoportal or data.europe.eu exist and offer the appropriate access tools, the majority of the data can only be obtained by digging through often untranslatable national websites or contacting the appropriate responsible staff directly via email. From a total of 16 countries (Austria, Belgium, Germany, Denmark, Estonia, Spain, France, Croatia, Lithuania, Latvia, Netherlands, Portugal, Romania, Sweden, Slovenia, Slovakia) that have released the data, only one used INSPIRE Geoportal, two published on data.europe.eu, three gave out the data via eMail, and the remaining states had the data only available on their respective websites. Some countries, including Spain, Belgium, and Germany, do not release the data on a national level; to get access to the information, one must visit an additional website for each region.

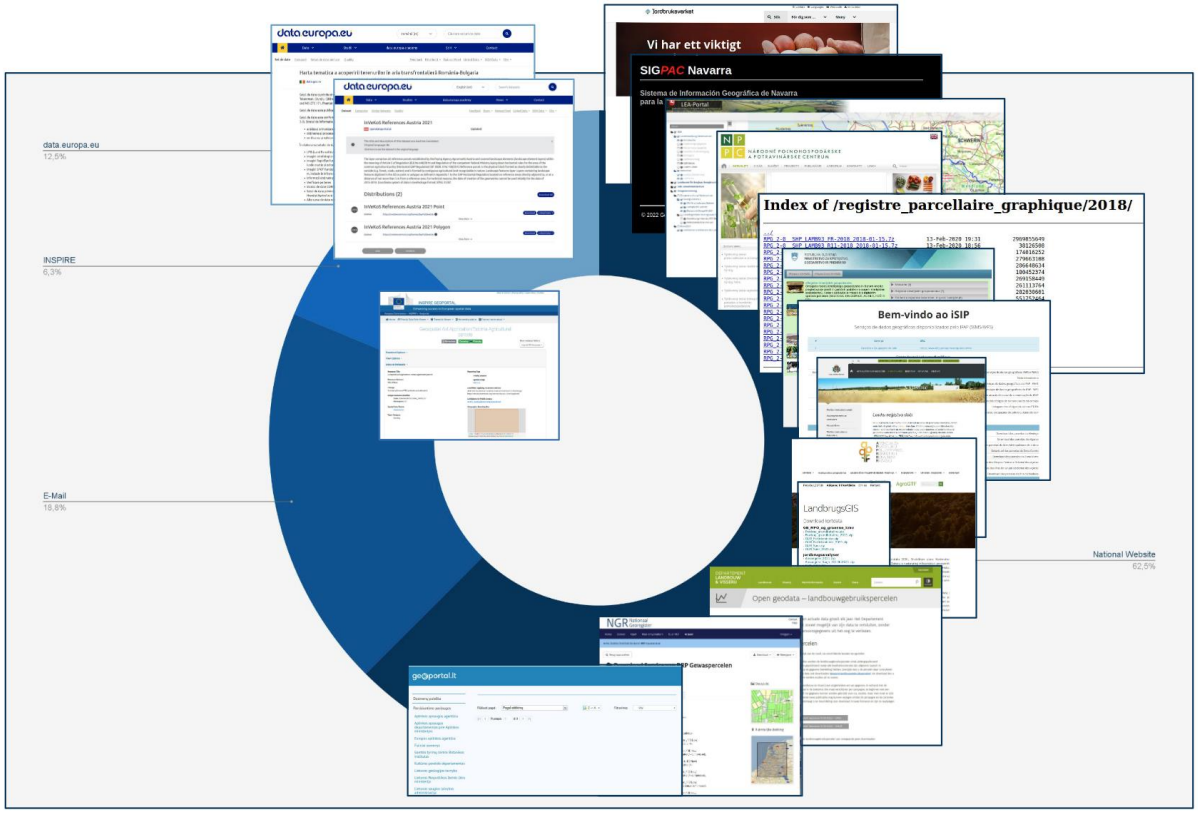


Figure 5: Heterogeneities of data release within Europe, © Schneider, Tech Univ Munich

Finally, when all available data is collected, the next interoperability problem arises: Published crop data does not follow a European standard and is therefore not comparable in any way. As a consequent step, harmonising the data is required. In the EuroCrops dataset, we introduced a new Hierarchical Crop and Agriculture Taxonomy (HCAT), which maps all country-dependent taxonomies into one unified schema, see table 1. Due to its hierarchical structure, any granularity in which the crop data might be declared, can be reflected and stored.

original_name	translated_name	HCAT2_Code	HCAT2_name
Ganyklos-pievos virð 5m.	Pastures-meadows over 5m.	3302000000	pasture_meadow_grassland_grass
Þieminiai javai	Winter cereals	3301011501	winter_unspecified_cereals
Ganyklos-pievos iki 5m.	Pastures-meadows up to 5m.	3302000000	pasture_meadow_grassland_grass
Vasariniai javai	Spring cereals	3301011502	spring_unspecified_cereals
Kita ariama þemë	Other arable land	3301990000	other_arable_land_crops
Rapsai	Rape	3301060400	rapeseed_rape

Table 1: Harmonised Lithuanian data, 2021. © EuroCrops

By now, for each of the 16 countries, the project has released one year of harmonised agricultural data with the outlook of temporarily extending it within the next year. The dataset, along with additional information, is available on https://github.com/maja601/EuroCrops.

6 Geospatial Data on data.europa.eu

Antje Kügeler, Con Terra

Jiri Pilar, European Commission

The European Commission is following the European Data Strategy outlined in 2020, see figure 6. For open data, it means monitoring and enforcing the transposition of the Open Data Directive in all EU Member States.

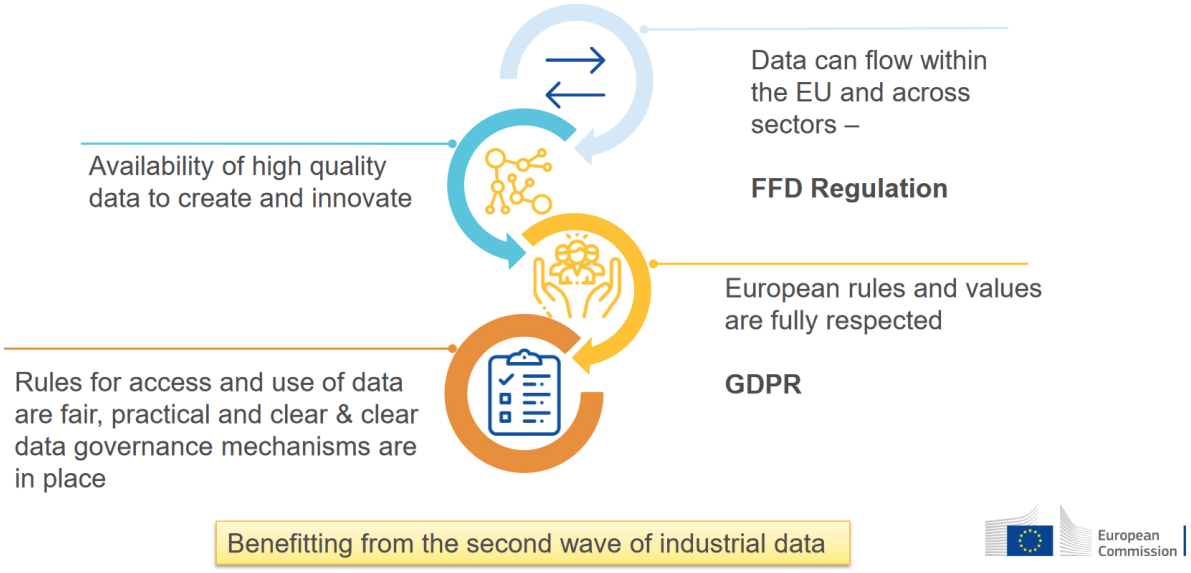


Figure 6: European strategy for Data. © European Commission

In 2022, the Commission identifies a list of 'high-value datasets', whose re-use can bring benefits for society, the environment, and the economy, see figure 7. Such data will have to be available for re-use for free, in machine-readable format, via APIs and, where relevant, as a bulk download. Geospatial data will represent a very important part of high-value datasets. In addition, two other upcoming pieces of legislation partly concerns public sector data: the Data Governance Act and the Data Act.

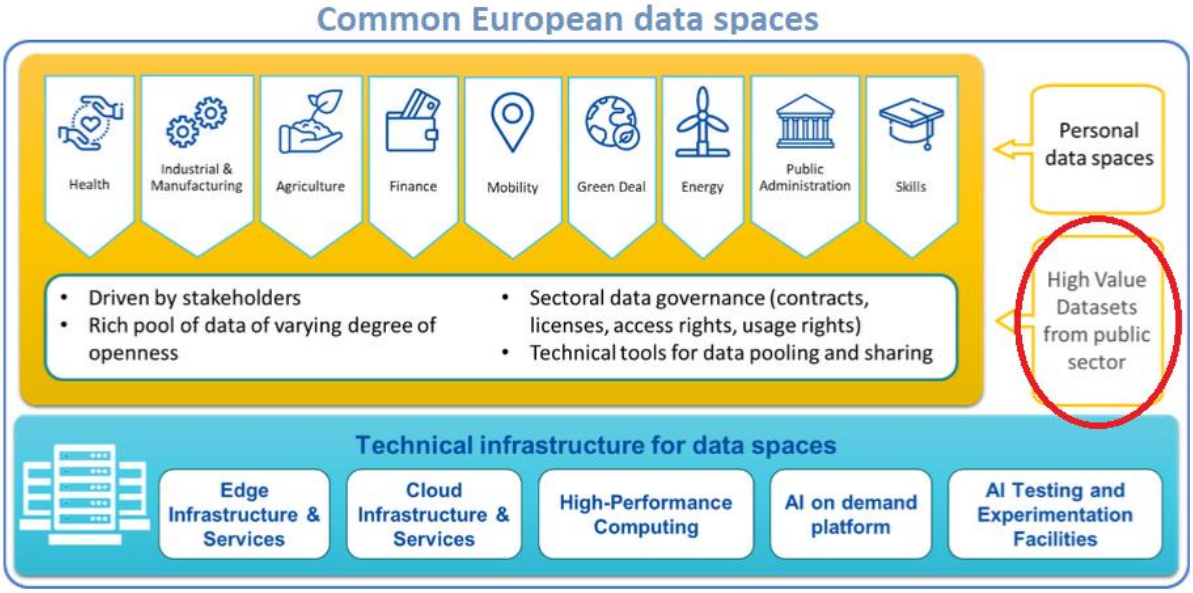


Figure 7: High Value Datasets and Common European data spaces. © European Commission.

In this context, the strategic objective of the official portal for European data, data.europa.eu, is to improve accessibility, promote the reuse of public sector information and generally increase the value of open data. In 2021 the two previous Open Data Portals "European Union Open Data Portal" and "European Data Portal" joined forces to provide a single point of access to open data from Europe. It provides access to open data from international bodies, EU organisations, as well as national and regional sources.

This is done by collecting the metadata of public data made available across Europe in a process called "harvesting". Metadata is harvested both from open data portals and geodata portals and is made available on data.europa.eu. Currently data.europa.eu provides access to descriptions of more than 1 million datasets from 173 catalogues and 36 countries.

With respect to geodata, many of the geoportals harvested by data.europa.eu are INSPIRE geocatalogues (or "INSPIRE discovery services"), which are also harvested by the European INSPIRE geoportal. There are also geoportals being harvested by Data.europa.eu that include no INSPIRE metadata or geoportals that include both INSPIRE metadata and metadata on resources not relevant for INSPIRE. During the geo-harvesting process data.europa.eu maps the INSPIRE metadata to GeoDCAT-AP. Not all mapped INSPIRE metadata elements are visible in the front end, to avoid overwhelming users with information, but they all are accessible via the SPARQL endpoint which stores the full result of the INSPIRE to DCAT-AP transformation.

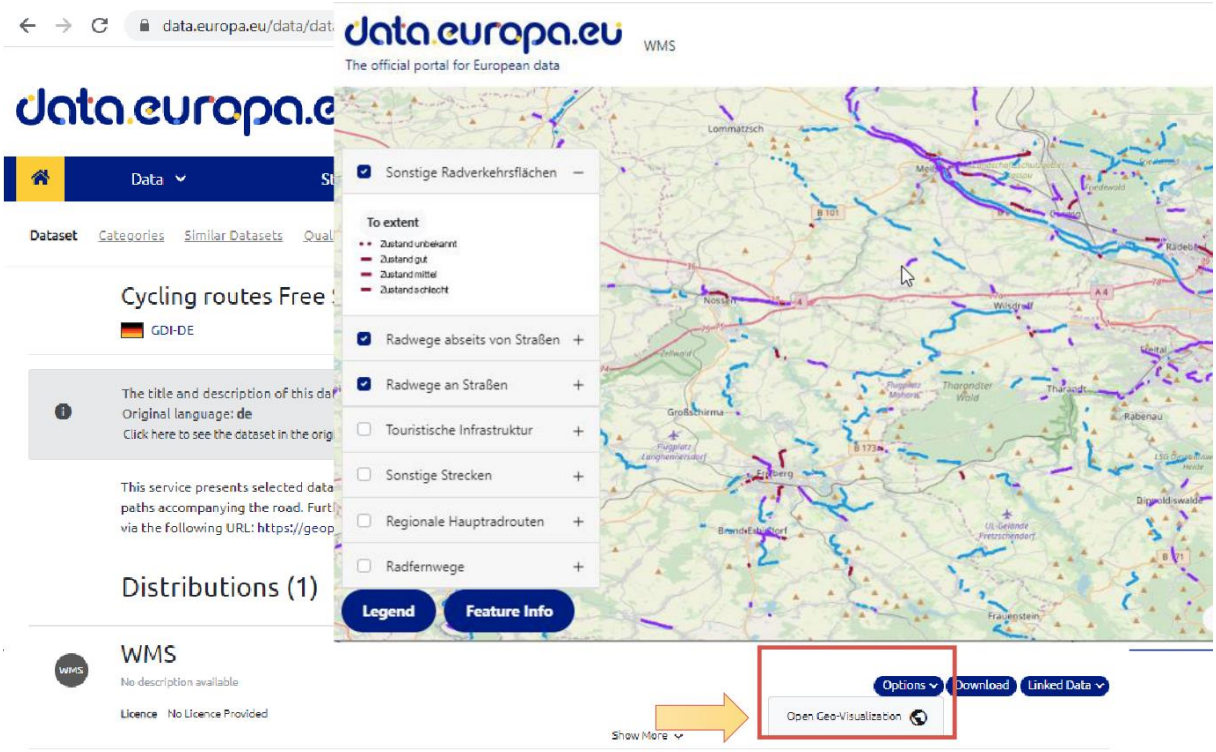


Figure 8: Example of display of a data source on data.europa.eu: description and preview. © Kügeler, Con Terra

To facilitate public data sharing it is possible to search across countries and particular thematic domains. The titles and descriptions of the described datasets are translated into the different European languages after the harvesting to allow a multilingual search. The users can discover data from different countries even if they do not speak the languages the data is described in.

Some popular geospatial content (Web Map Services – WMS – and GeoJSON) can be previewed in the map viewer that is part of data.europa.eu, see figure 8.

Apart from the different options to search for, browse and discover data on data.europa.eu, it is important to keep in mind that some users may not even be aware of the portal's existence and its content. These users will be more likely to use mainstream search engines than go to specialised data portals. To enable discoverability outside of data.europa.eu special information (schema.org in JSON-LD encoding) is included in the html pages. This is not visible to human users but makes it easier for search engines to index the content, allowing datasets to be found with mainstream search engines.

7 The Copernicus Data Stores implemented by ECMWF

Ángel Lopez, Copernicus

Copernicus is the European Union earth observation program and combines different kinds of satellites. The European Centre for Medium-Range Weather Forecasts (ECMWF) has been mandated to implement and run specific Copernicus services based on these satellites data : Climate Data Stores (CDS) and Atmosphere Data Store (ADS), https://cds.climate.copernicus.eu/, https://ads.atmosphere.copernicus.eu/. CDS and ADS share the same infrastructure operated and

maintained by a DevOps team. The design is flexible and adapts to different technologies currently adopted by users to search datasets, as well as an APIto access the datasets, https://cds.climate.copernicus.eu/api-how-to, see figure 9.

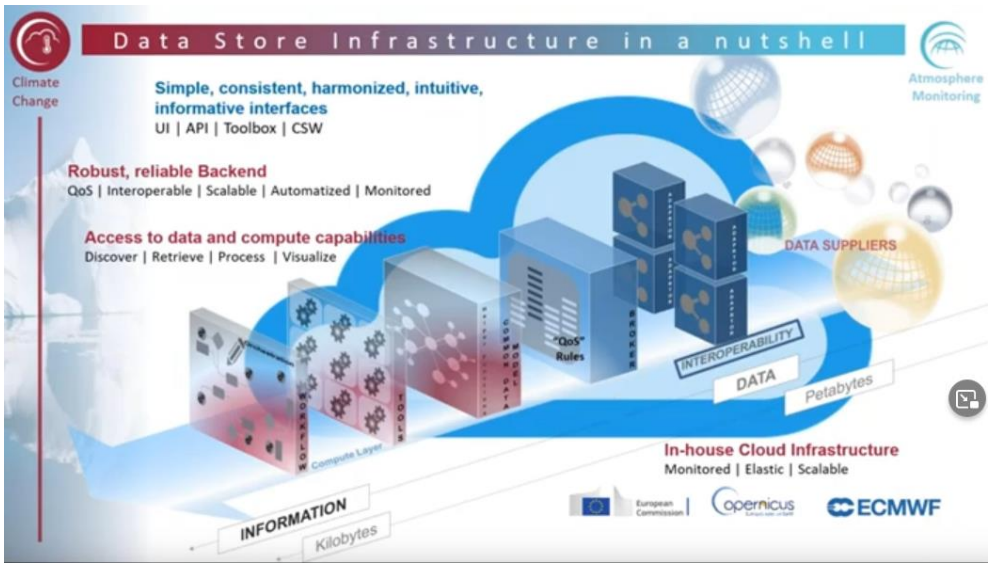


Figure 9: Overall architecture and technologies considered in the data store infrastructure, © Copernicus

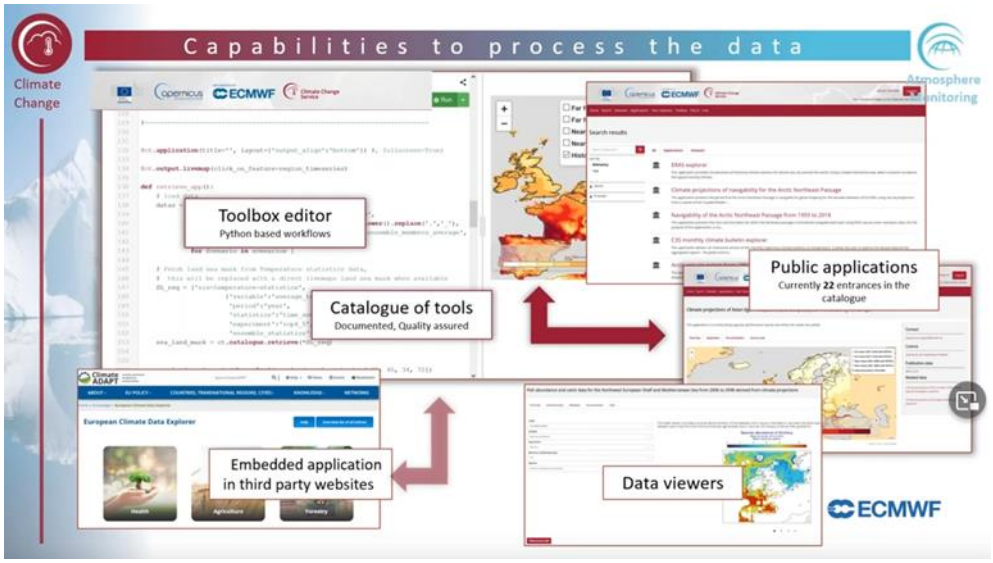


Figure 10: Additional resources provided by the CDS to process data, © Copernicus.

Besides, the project also embeds resources to process the data, tools and catalogues of tools as illustrated in figure 10.

Current perspective is to bring computation closer to the data, with an advanced use of cloud computing, and to achieve a more Streamlined Climate Data and Information Flow using Copernicus Data and Information Access Service (DIAS), https://www.wekeo.eu/.

8 INSPIRE Geoportal – Towards the new EU data ecosystem

Jordi Escriu, Joint Research Center

The European INSPIRE Directive entered into force in 2007. It engages Member States to contribute to a European Union spatial data infrastructure, whose main purpose is to serve EU environmental policies as well as EU policies or activities which may have an impact on the environment. Member States shall contribute by facilitating public access to their public sector spatial information. This multifaceted initiative is organised through a framework covering the legal, organisational and technical data governance aspects. Today it is one of the biggest geospatial data sharing initiatives in the World with more than 7000 data providers involved. After 15 years of implementation, the goals of the Directive have been partially achieved, and some of its challenges still remain. Concretely, heterogeneity of implementations across Europe call for additional brokering and pan-European coverage is yet to be achieved.

INSPIRE is currently immersed in a deep process of adaptation to better serve the digital and data priorities of the European Commission. In the context of the Digital EU priority (a Europe fit for the Digital Age) and the European Strategy for Data, INSPIRE should evolve into the so-called European Green Deal data space, illustrated in figure 11. As a driver to achieve this evolution, a process of evaluation and revision of the INSPIRE Directive is currently being undertaken, including all the lessons learned from its implementation, which constitute an essential contribution from the public sector to the envisaged sectoral European data spaces, in particular to the Green Deal, as analysed by the JRC, Geonovum and DG ENV [Kotsev et al. 2021].

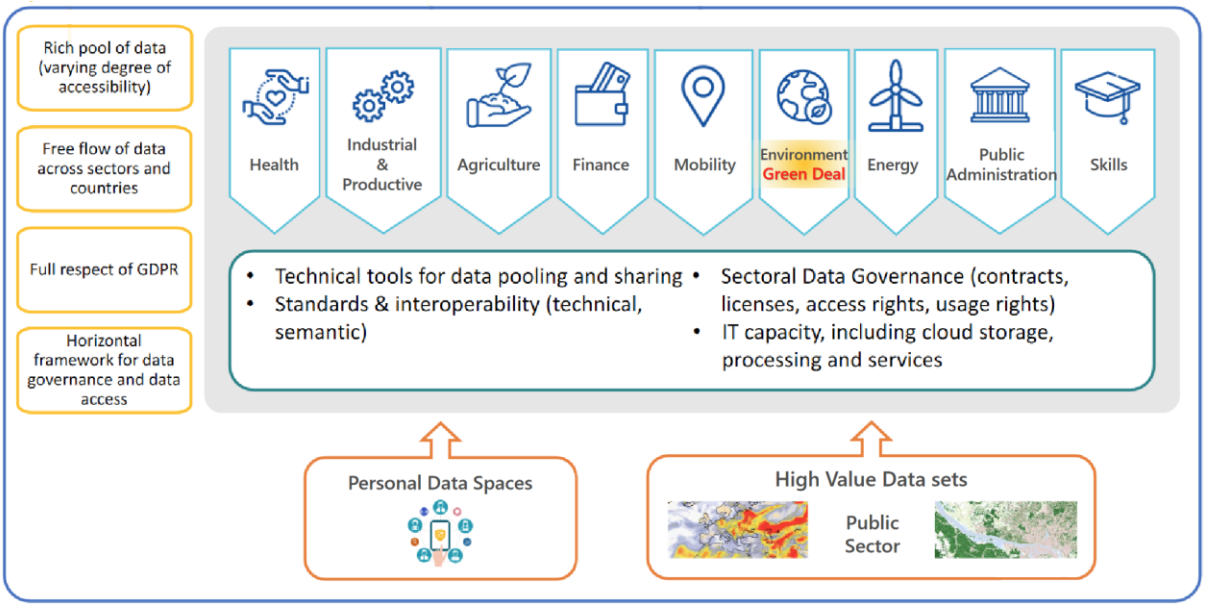


Figure 11: Sectoral European data spaces, © European Commission.

The vision developed after this analysis considers:

- the evolution of INSPIRE to the concept of data ecosystem.
- broadening its scope to new sectors (not only the public sector, but also the private sector/businesses, and academia).
- engaging with new communities, such as developers, and end users.

- widening the range of applications and use cases.
- making the INSPIRE framework simpler, more flexible and agile.
- lowering the knowledge entry-level for implementing and/or using data.
- reusing well-adopted standards and technologies that proved to be in use.

The INSPIRE Geoportal (https://inspire-geoportal.ec.europa.eu), which is the main access point to INSPIRE Infrastructure tools and resources, is now being revamped, including a new user interface (frontend) which will include the possibility to browse and filter INSPIRE datasets according to High-Value Datasets categories envisaged in the Open Data Directive, as well as a new catalogue system (backend) based on GeoNetwork open source, using an architecture migrated to a cloud environment.

Moreover, the JRC is researching to establish the technical and non-technical requirements for the evolution to European data spaces. The outcomes are being included in a 'Data Spaces Cookbook', co-created and validated with different European Commission services, but also other sub-products have been produced (an online living document, an interactive component / chatbot for Q&A purposes) and are already available.

9 GeoNetwork opensource - New developments and plans

Jeroen Ticheler, GeoCat

GeoNetwork is an open source and standards-based technology to manage and expose spatially referenced resources. The tool started in 2000 at the Food and Agriculture Organisation of the United Nations (FAO) only counting with one developer. Now it is supported and used by an active and global community of developers and users. There are around 1000 software downloads per month and its technology is used by 85% of the current national INSPIRE catalogues (discovery end-points). Metadata edition and multilingual metadata support was the focus of version 3 of the software.

The latest version of the tool, v4, focuses on the searching functionality. Its architecture is displayed in figure 12. It integrates spatial search within the index, and is about 10 times faster than the old searching mechanisms. Both, index and search components, run separately from the rest of the application, which improves performance and scalability.

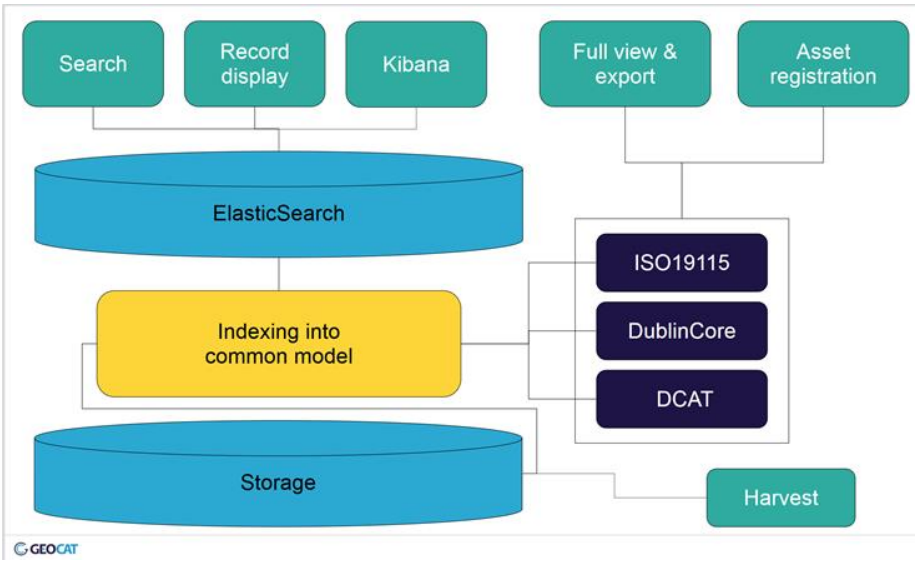


Figure 12: GeoNetwork v4 architecture, © Ticheler, GeoCat

With regards to standards, OGC API - Records is an alternative to OGC Catalogue Service for the Web (CSW), and it is implemented on top of GeoNetwork v4. This will still require some refactoring in the future, since the standard is not approved yet and might evolve.

Additionally, GeoNetwork v4 adopts a micro-services architecture as shown in figure 13. This brings benefits, for example the ability to run parallel background tasks, improved resource management, the possibility to mix and match components, scalability and happy DevOps. It required and still requires a modification of the monolithic application GeoNetwork used to be.

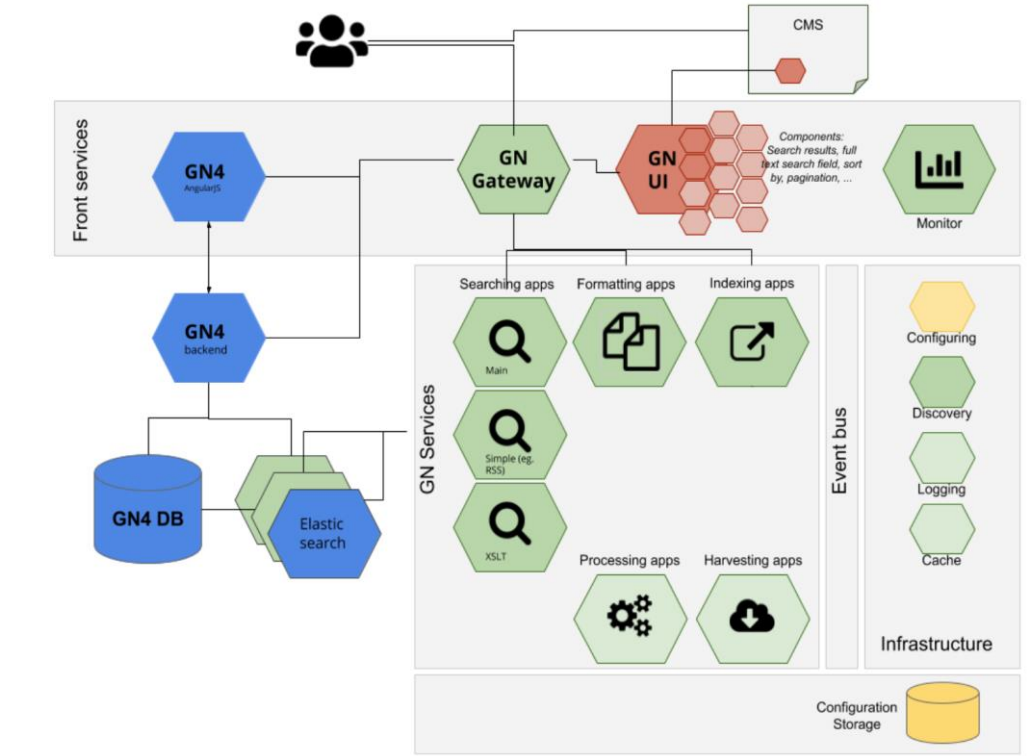


Figure 13: Microservices in GeoNetwork v4, © Ticheler, GeoCat

New developments have been made for the implementation of the revamped INSPIRE Geoportal. This 'Super Harvester' is based on three new components: 1) an improved and high-performance harvester, 2) a link checker validating the real accessibility of data and services , and 3) an ingestion engine to publish the results obtained from the harvesting processes, see figure 14.

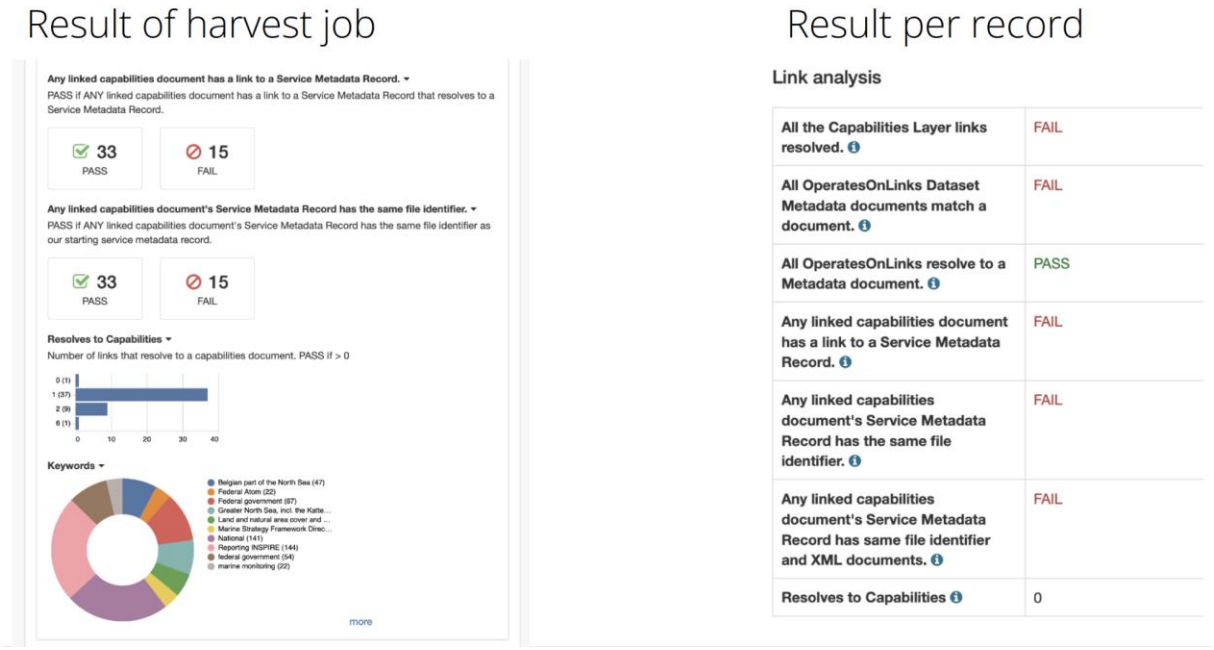


Figure 14: Example of results from the link checker component, © Ticheler, GeoCat.

Other developments relate for example to the open data support, amongst others through a (geo) DCAT-AP schema-plugin and the related metadata editor.

10 Geospatial metadata on the (Semantic) Web. Lessons learnt in GeoDCAT AP

Andrea Perego, European Parliament

DCAT-AP is the de facto standard for metadata interchange across European data catalogues, based upon the W3C Data Catalogue Vocabulary (DCAT). GeoDCAT-AP, https://semiceu.github.io/GeoDCAT-AP/releases/, is a geospatial extension to DCAT-AP developed in 2015 by a working group involving experts and stakeholders from EU Member States, chartered by the EU ISA Programme (now Interoperable Europe). The purpose is to increase the discoverability of geodata outside the GI community by adopting a more generic standard and extending this standard to adapt to the specificities of the GI domain, as shown in figure 15.

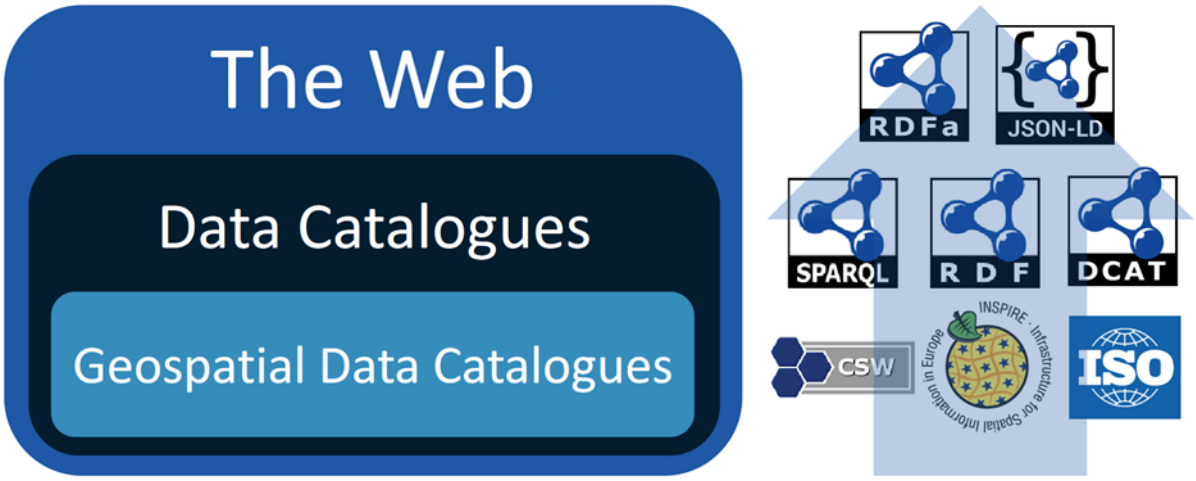


Figure 15: Improving the visibility of Geodata on the Web, © Perego, European Parliament.

It targets the definition of a harmonised RDF transformation & representation of geospatial metadata, https://github.com/SEMICeu/iso-19139-to-dcat-ap, the sharing of geospatial metadata with general-purpose data catalogues, and support of cross-platform discovery of and access to geospatial data.

A demonstrator (http://geodcat-ap.semic.eu/api/) was developed to :

- provide a working example on how GeoDCAT-AP can be supported without changing the underlying infrastructure
- show how to enable standard HTTP functionalities in CSWs, as HTTP content negotiation
- show how to increase visibility on the Web for geospatial metadata, using standards as HTML+RDFa

Lessons learned after GeoDCAT-AP v1 (2015) were identified. There is a lack of common practices to specify some geospatial information in RDF. The use of global & persistent identifiers in geospatial metadata is too limited. Version 2 of GeoDCAT-AP was released in 2020 and addresses these limits together with the latest versions of the INSPIRE Technical Guidelines for identifiers.

11 OGC API features – building blocks to create, modify, and query geospatial vector data on the Web

Jari Reini, National Land Survey Finland

API development at National Land Survey in Finland is driven by several considerations. NLS Finland's data is extensive and constantly updated. Timeliness is a key factor for the user. The users need the data as part of their own process and information systems. Last, with the issue of the EU's Open Data Directive, NLS data are among HVD's (High Value Datasets).

In that context, the Spatial Data on the Web Best Practices are :

- to use globally unique persistent HTTP URIs for Spatial Things
- to make your spatial data indexable by search engines

- to use spatial data encodings that match your target audience
- to expose spatial data through 'convenience APIs'
- to include spatial metadata in dataset metadata

Additional recommendations are listed as being FAIR: Findable (through catalogues and metadata), Accessible (by using an open and standardised protocol), Interoperable (using standards for content encoding) and Reusable (by providing licence information and data product specifications).

In that context, the OGC APIs are new standards that take advantage of modern web development practices to define resource-centric APIs: Features, Maps, Records, Processes, Coverage, Environment Data Retrieval. They build upon the legacy of the OGC Web Service standards (WFS, WMS, WCS, WPS, etc.) and are being constructed as "building blocks" to support download, updating, processing, view and discovery.

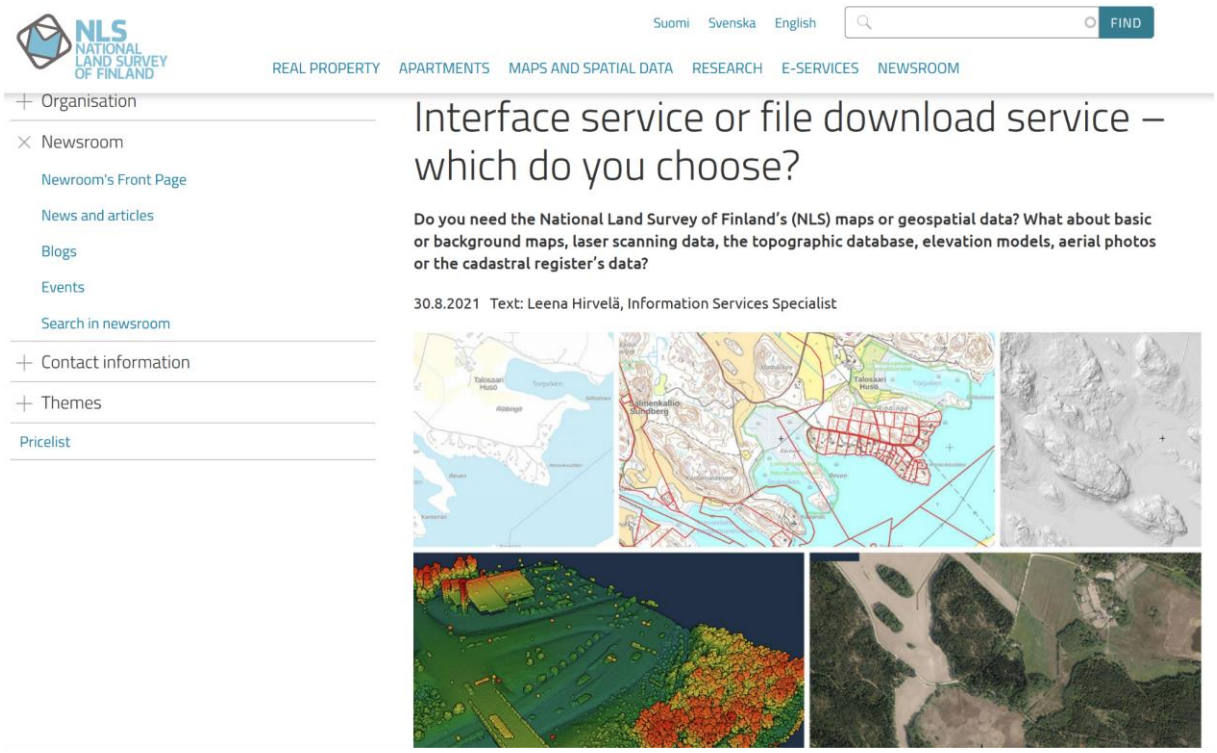


Figure 16: Implementation of OGC APIs on NLS portal, www.nls.fi

NLS has adopted OGC APIs for its open data to be published online : background maps, spatial data (such as place names, roads, buildings, addresses, water bodies and other terrain patterns, elevation ratios and administrative boundaries), real estate information, residential and commercial property information. The portal is shown in figure 16.

12 Linking Authoritative (Government) data with Community data (e.g. Wikidata)

Erwin Folmer, University Twente

Data about buildings is stored in different kinds of datasets; including government data, commercial data, and community data. Most of these datasets have been set up as silos. There is a lot of potential in connecting all those isolated data containers. It makes more value for the end user, by t improving the quality by redundancy and complementarity of the sources. In that context, a lot of effort has been put on connecting data from different origins within government sources into a Knowledge Graph, illustrated in figure 17.

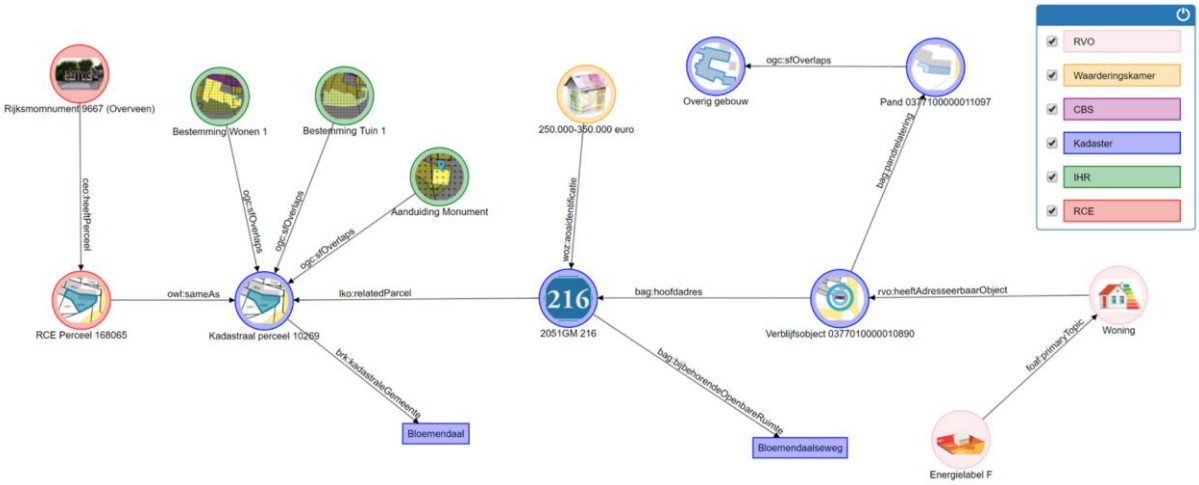


Figure 17: Interconnection of government data related to a building with the Kadaster Knowledge Graph, © Folmer, Uni Twente

Next step is to connect with other data, to reach the potential of linking to community data.

Kadaster investigated the creation of links between Wikidata and the Dutch Cadaster, as shown in figure 18. While Kadaster created 50 links in 2020, over 20 000 links have been created by the community, to associate buildings with community data like, for example, images.

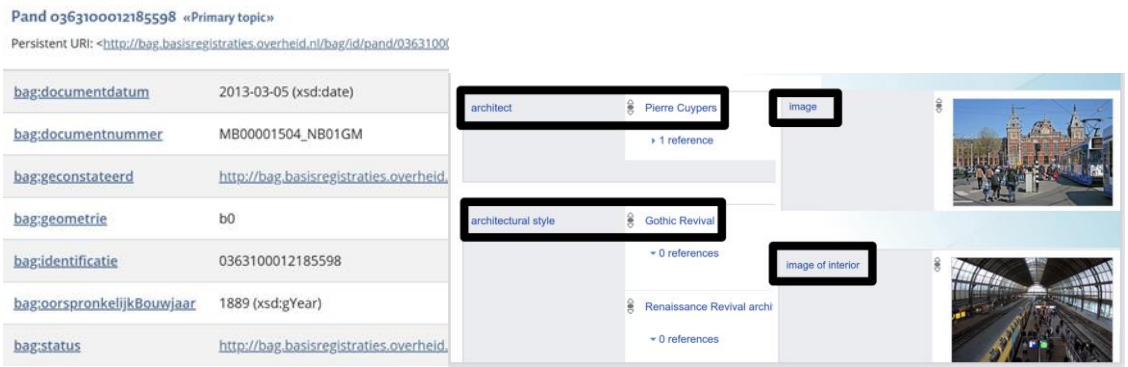


Figure 18: The same building depicted in the Base Registry (left) and in Wikidata (right), © Folmer, Uni Twente

Kadaster recently published several ways to access the integrated data of the Dutch government through Linked Data technology, starting from publishing individual features within its databases.

As the creation and maintenance of data like the Base Registry for Topography (BRT) is expensive, a relevant question is the enrichment of buildings data with an external open data source (e.g. OpenStreetMap). The project explored which building types and OSM data will contribute to the highest value creation. It automated the OSM data ingress process, created a geometric match and transformed OSM to Linked Data (including geometric links) and connects to the existing BRT Linked Data set, as illustrated in figure 19.

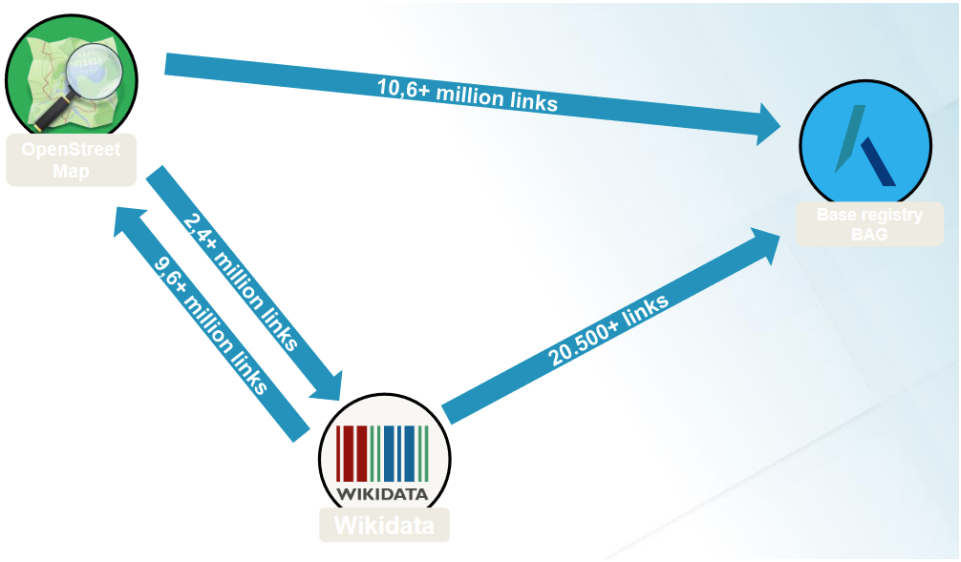


Figure 19: Linking Base registry for Topography (BRT) with Wikidata and OSM,
© Folmer, Uni Twente

This process is not straightforward and the project found many semantics and quality issues. Several challenges need to be solved to achieve quicker and effective output. So to say, there is a big win but not a quick win.

13 Beyond SDI? Becoming cloud native?

Ed Parsons, Google

Cloud technology calls for an upgrade of SDI infrastructure to reach cloud native SDIs. Access and usability are the two pillars of successful spatial data publishing. Metadata harvesting is not enough. Data must be discoverable by automated systems and data must itself be directly accessible and linkable on the Internet. Big volumes of spatial data, especially earth observation (EA) data, requires streamable data formats.

A starting point for dataset search from Google perspective is to follow 4 steps illustrated figure 20: add markup to the dataset pages, verify the markup, and submit a site map.

Another approach is to enhance data discovery by browsing.

Last, data must be steamable to manage large volumes.

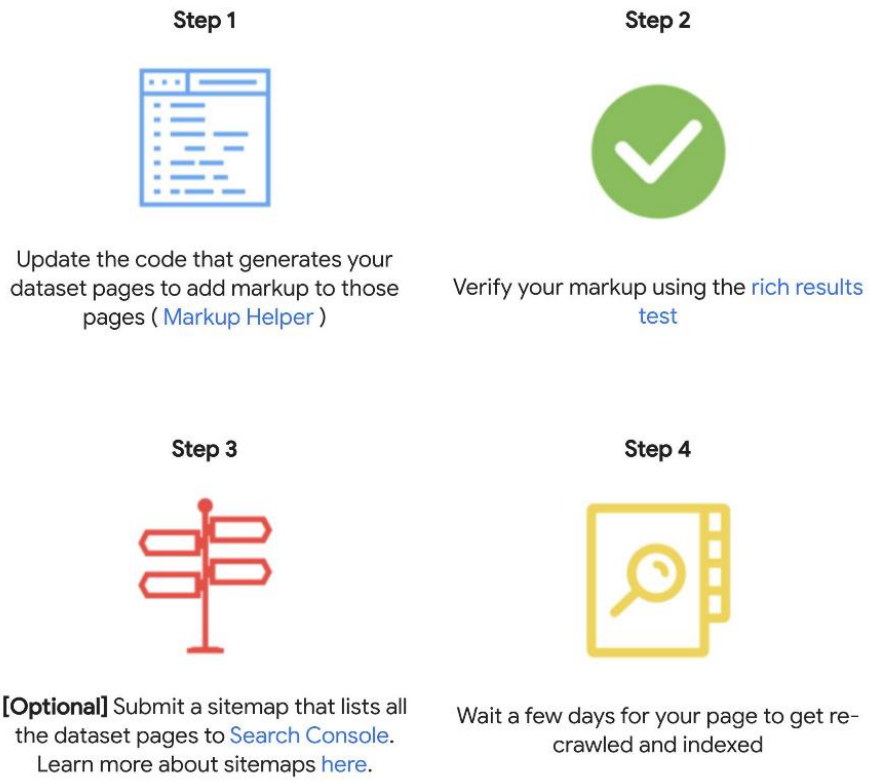


Figure 20: Preparing dataset to be searchable by engines, © Parson, Google.

14 ESRI vision: Maximising discoverability and access in a system-of-systems

Jill Saligoe-Simmel, ESRI

How we think of traditional Spatial Data Infrastructure (SDI) is evolving. Because, of course, technology is not static but is constantly evolving. Geospatial data supply chains are shifting from a hierarchical to circular data economy, with digital communities of practice consisting of data providers, coordinating bodies, and data consumers. Today, organisations come together to collaborate and share in digital ecosystems, a truly integrated geospatial infrastructure, as illustrated in figure 21. New patterns for modern SDI are shifting paradigms from search and discovery to integration and use. How do we maximise geodata discoverability and access in this system of systems?

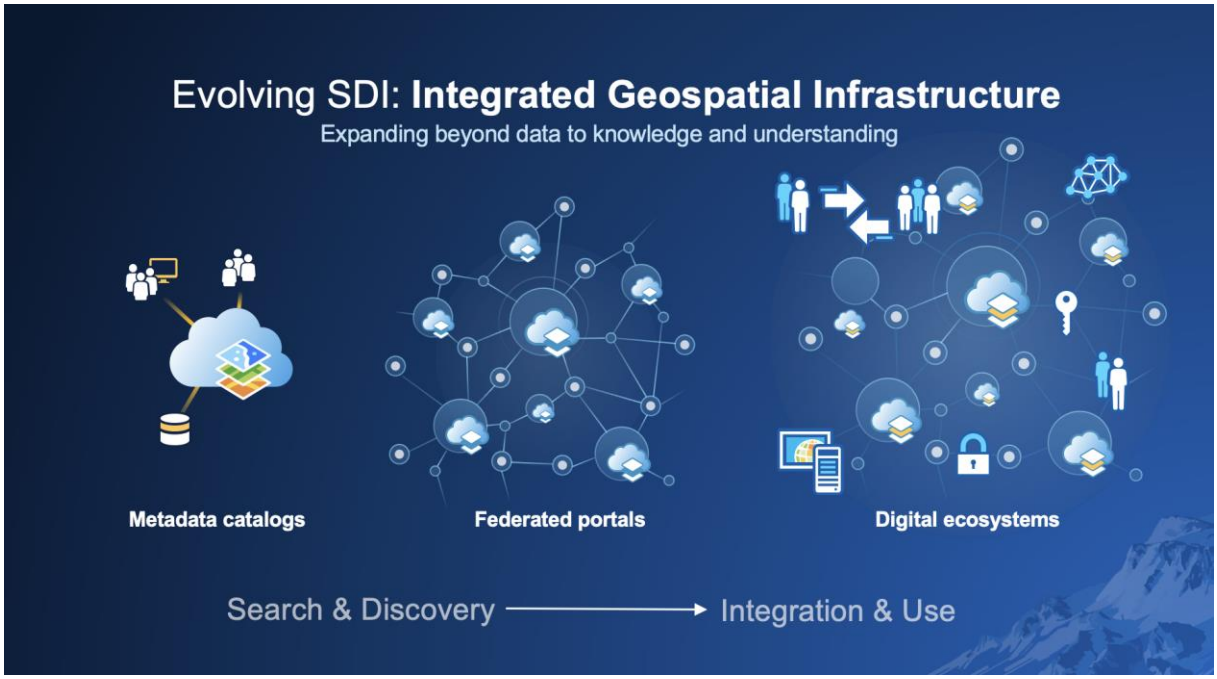


Figure 21: Integrated geospatial infrastructure, © Saligoe-Simmel, ESRI

Activities that promote discoverability and access must span the end-to-end workflow of providers who produce, publish, collaborate, share, and (re)use data, see figure 22. For example, data providers at all levels of government and across all industries produce and manage data and metadata using domain-relevant community and international standards. These organisations then publish data as interoperable web services and APIs using open standards and specifications, ensure their content has proper metadata, and carry appropriate licensing.

Figure 22: Support for discovery and access spans end-to-end workflows, © Saligoe-Simmel, ESRI

As illustrated in figure 23, cloud-native platforms enable frictionless sharing and collaboration among trusted partners who work together to address shared challenges. Content is shared with relevant audiences as open or private data, based on the permission granted by the data owner, and described with clear licence terms of use. Catalogues index the streaming web services across multiple organisations without duplication, search engine optimise (SEO) the content, and federate to open data portals using standards like INSPIRE Metadata and automatic transformation to DCAT-AP. These

activities, combined, are consistent with the FAIR data principles – Findable, Accessible, Interoperable, and Reusable [Saligoe-Simmel et al. 2022].

Figure 23: Collaboration in cloud-native platforms, © Saligoe-Simmel, ESRI

The FAIR principles come to life when curated content is shared through ArcGIS Hub. ArcGIS Hub is Esri's next-generation community engagement and collaboration platform supporting content delivery, including open and secure data, web services, documents, maps, tools and apps, and their closely coupled metadata. Unlike traditional metadata catalogues that harvest standalone metadata, Hub dynamically indexes data and tightly coupled metadata through a group sharing mechanism where people and content participate in open and secure collaboration spaces. Collaboration spaces are easily interconnected, creating a network of Hubs in an expanding digital ecosystem where sharing and collaboration happen at scale.

Digital ecosystems place a renewed emphasis on data consumers. Data producers realise increased return on investment as shared web services are used and reused. In addition to web services, sharing through maps and apps, where data are visualised and explored, expands data access to broad audiences who may be unfamiliar with GIS. As a result, value is unlocked as data are accessed and used to drive decision-making and innovation.

Today, this vision is materialising among thousands of organisations worldwide. Tools are helping to maximise discoverability and access through user interfaces, automation, workflows, infrastructure, interoperability, and compliance in ways that are easy to implement. As a result, an interconnected system of systems is quickly emerging – a global digital ecosystem – enabled by geospatial infrastructure that empowers organisations to collaborate and be more effective than ever.

15 Rediscovering Spatial Data Discoverability

Jordi Escriu, Joint Research Center

During the last INSPIRE Conference 2021, a session on Spatial Data Discoverability trends was offered to the community. Metadata catalogues are evolving at a very fast and dynamic pace. Member States are developing cataloguing solutions that encompass both INSPIRE/geospatial as well as Open Data domains.

The new developments are pushing the evolution of software components, in particular open source ones, making metadata available in multiple national and European infrastructures through different standards and rules.

The OGC is finalising the new OGC - API Records standard, which will highly simplify the discovery, access and management of metadata on the web. In addition, the SpatioTemporal Asset Catalogues (STAC) has been identified as an emergent de-facto standard for publishing geospatial data catalogues online. Finally, GeoDCAT-AP is being highly used for making spatial data publicly and consistently accessible in both, the Open Data and the geospatial domains.

All these practices have been recently identified by the INSPIRE community in the fields of Spatial Data Discoverability. As a result, they have been proposed as 'INSPIRE Good Practices', which is the mechanism to analyse new trends and working standards available in the market with the view to facilitate its adoption in INSPIRE, as part of its evolution process (https://inspire.ec.europa.eu/portfolio/good-practice-library).

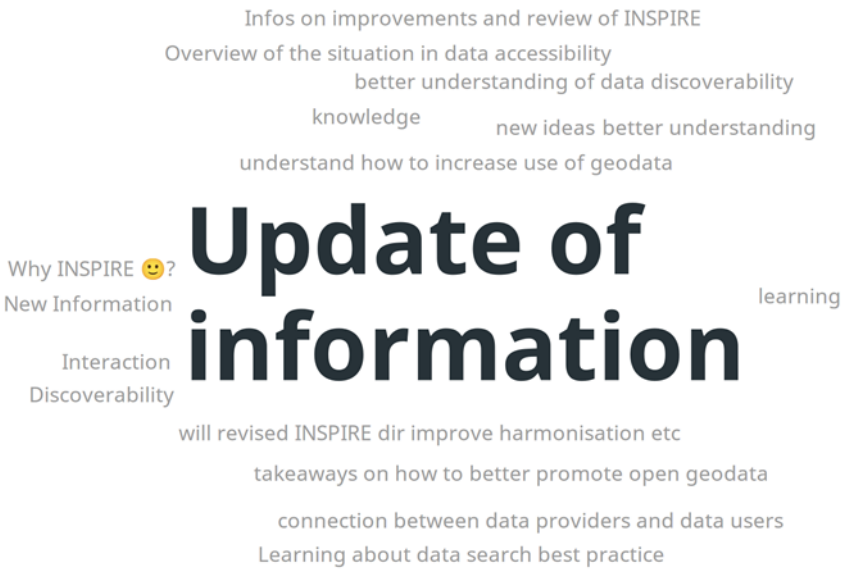
As illustration, OGC API - Records has been already endorsed as a INSPIRE Good Practice candidate to simplify not only the discovery, access and management of metadata on the web, but also to make such resources automatically discoverable through search engines.

As final reflections on data discovery, future metadata models should be storing the information in a standard-agnostic way, and metadata catalogue tools be based on the emerging standards mentioned hereabove. Besides, users expect simple, understandable communication in user interfaces (frontends), such as searching based on Natural Language Processing (NLP), data filtering, and levelled exploration of results. Metadata results should be automatically translated to the user's language and adapted to the user's knowledge. Finally, the frontend shall not be necessarily coupled to the SDI geoportal / catalogue.

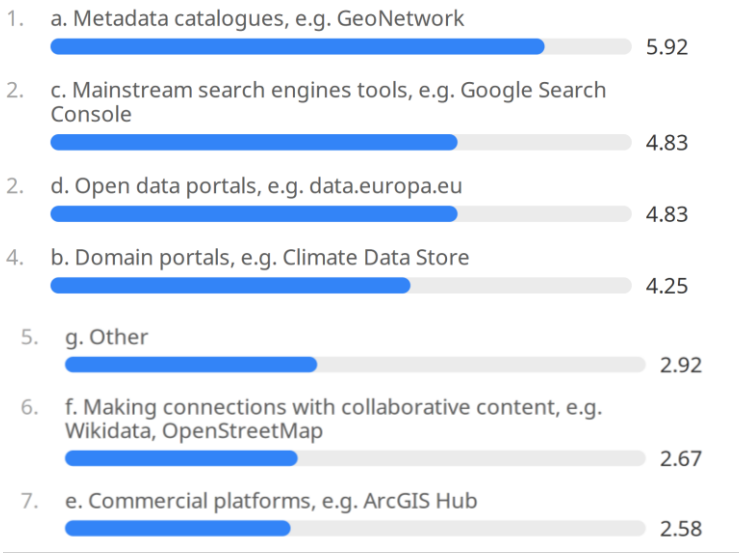
16 Wrap up poll

The last session was dedicated to a live poll between participants who were willing to participate. The first question was answered by 21 participants. The latter were answered by an average of 12 participants.

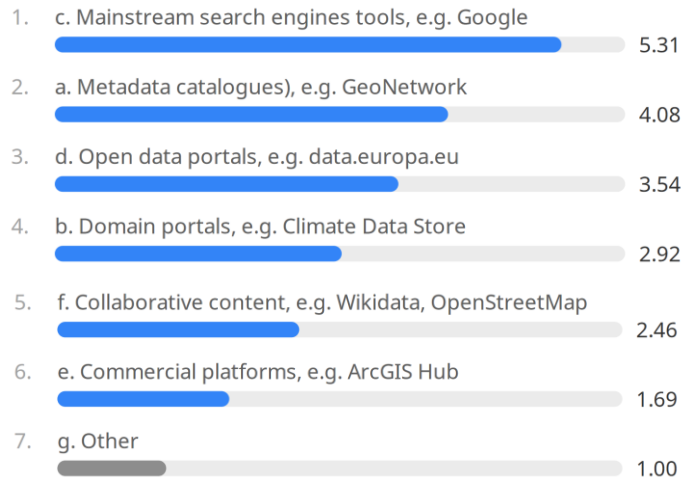
Question: What are you looking forward to from this workshop? (free text answers)



Question: As a data provider, when publishing information about spatial dataset on the Internet, please prioritise tools you use to improve the dataset discoverability. Put the most used tool at the top, and the least used at the bottom.



Question: As a data user, when searching for information about spatial dataset on the Internet, please prioritise the tools you use to find relevant data. Put the most used tool at the top, and the least used at the bottom.



Question: Do you use other tools than those mentioned in the previous question to find spatial data?

7 answers No.

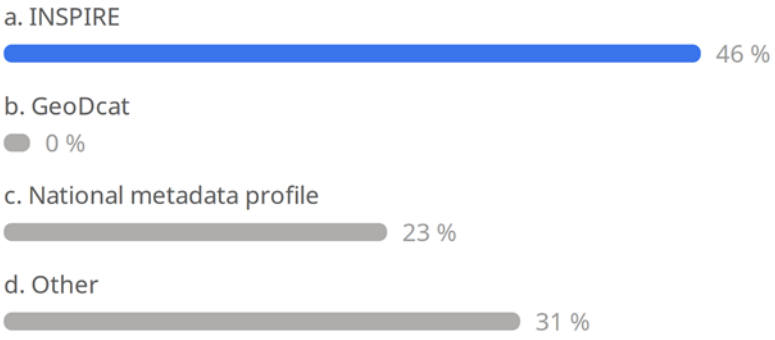
Other: local authority data portal, national geoportal, kohesio, zenodo, local contacts

Question: What insights do you require from discoverability tool providers regarding their tools? (free text answers)

- Completeness of coverage
- Easy-to-use editor, with automatic aids. Ability to transform metadata on-the-fly and serve the information in a wide range of standards. Bridging the gap across different domains (eg Geospatial vs Open data). Metadata exposed through the use of geospatial discovery APIs. Automatically discoverable through Web search engines. Discovery service (backend) in the cloud.
- Support of standards from the 19100 series as well as OGC. In addition to this we all look on how to integrate with our production system
- How to ensure that the published data is discoverable by a search engine, is compatible with a national portal, and that published metadata can be easily updated.
- Transparency, metadata
- Detailed manuals and tutorials
- How they rank (Google dataset search).

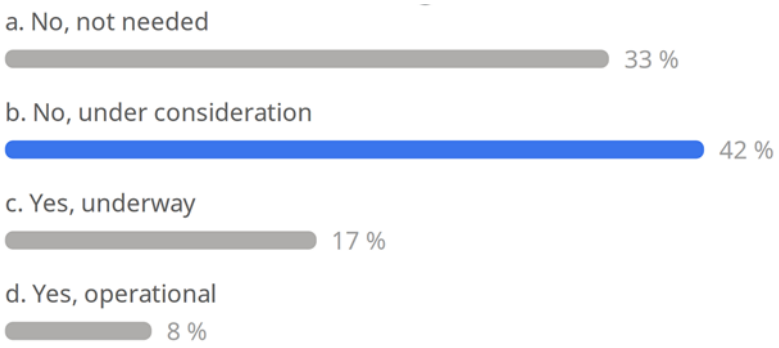
Question: Do you think that data providers are doing enough to improve their data discoverability?

Question: What is the favourite standard used to profile the metadata in your organisation?

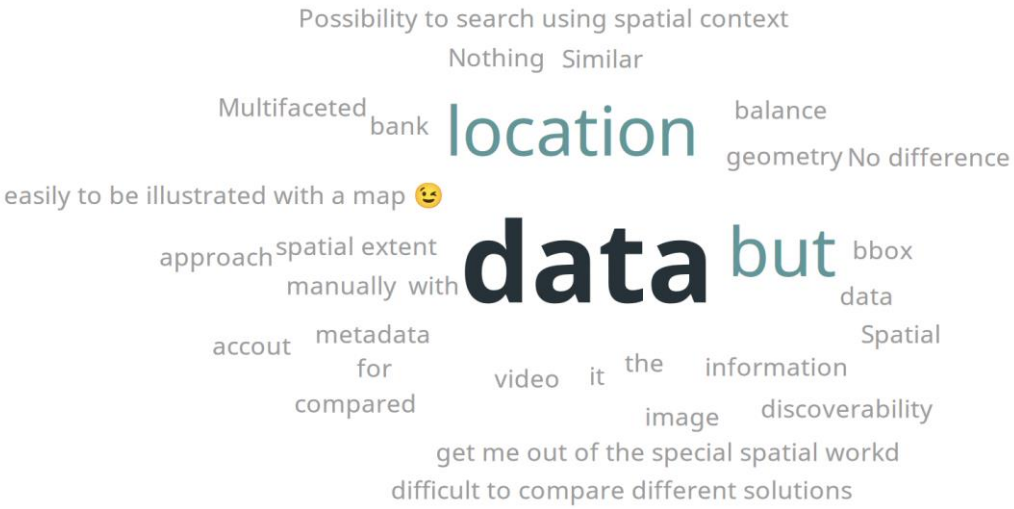


Question: What are the key elements you find the most relevant in the metadata discovery content or you would like to get, e.g. links to services, indication whether data is open, lineage, contact to data provider?

Question: What about the conversion between INSPIRE/(Geo)Dcat? Have you implemented specific tools for that purpose?



Question: What makes spatial data discoverability so different from other data discoverability (please provide keyword)?



Question: What would you demand of spatial data discoverability in order to generate a business value? (free text answers)

- Easiness of discoverability
- Being integrated in Web search engines. Simple, understandable communication with the user. Including NLP search, data filtering capabilities, levelled exploration of results. Automatic translation of metadata Discovery application not necessarily coupled to the SDI geoportal / catalogue.
- data which are usable not requiring XX days to get it working
- I don't know about business value but lots of missed data use and opportunities due to lack of discoverability
- End users must be taken into account - not only the developers but also the citizens/decision makers. They also need to know something about the quality of the data used e.g. for the COVID dashboard to see if they should trust the information (and hence need some basic data literacy to understand enough to know whether to trust what they see).
- open data
- For the data providers, a key factor is finding new users for the data. Therefore, I am investigating if discoverability returns a specific number of new users. For data users, a key aspect is the efficiency of search tools. Meaning how fast I find the data I need.

Question: Do you have any suggestions for follow-up activities?

- Hands-on workshop
- Have another workshop where we could have some more open discussion and interaction between the speakers and audience.
- Report will be of great interest. A session during next INSPIRE conference Provide feed-back to MIG-T. Difficulties of users in searching data to be widely communicated (e.g. by INSPIRE community) => encourage efforts from data providers, search engine
- Thanks everyone for a great workshop! For follow up as well as publishing all the videos and the slides and polls etc, could we maybe organise something similar with more of a developer perspective - i.e. to find out what do developers / analysts/ AI machine learning people who are using the data need to know? How do they feel about discovering data? Can they find what they need? Also publish a list of where funding is available for research into this topic.
- Hands on and practical sessions: Some tutorials about how to implement some of the things we've seen ... how do you build linked-data/rdf? How to query linked-data? Knowledge Graph? How to produce linked data? How to build a taxonomy etc. How to use various tools (from catalogue to cloud tools)
- Organise a follow-up event, inviting geospatial data providers, specifically asking them which geodata discoverability aspects would change after having attended this workshop.
- Geospatial Open Data - EUROGI good practices workshop in May
- Design case studies to benchmark solutions

17 Conclusion

To conclude this report, much work has been done and is on-going on the topic of (geo)data discoverability. Contributions stem from different actors: software developers, the INSPIRE community, local, national or European open data portals. Discoverability can also be improved by modifying the data themselves and using shared schemas and taxonomies to structure them.

Yet, despite these active contributions, progress is still needed. This progress will require the collaboration of different parties involved in the data life cycle and at least a common vision on what discoverability is. It is important to analyse what is discoverability nowadays in the information systems. In particular, different users must be considered, developers -of different components- and end users who make decisions or produce derived knowledge based on the data. Discoverability of data should not be a preliminary sequence in the development of applications but rather be intertwined with the development and exploitation of the data.

There are many technical solutions, possibly emerging, to improve the current situation, and neither existing solution is universal for every use case. In most cases, optimising geodata discoverability requires utilisation of different tools and technologies which complement each other. Tools combination maximises the chance of potential users to find data relevant to their needs.

Last but not least, the workshop evidenced the relevance of knowledge exchange on that topic and the willingness of participants to renew the experience, to get data providers more involved on this topic, to tackle more technical aspects and organise tutorials and hands-on sessions. After this broad scope, participants suggested to structure follow-up exchanges through community focuses.

Acknowledgement

The authors warmly thank Matina Fuentes, Patricia Sokacova, Alina Talipova and Carol Agius from Eurogeographics for their contribution to the workshop organisation and animation. We also thank Eric Bayers from NGI Belgium, Julián Hernández Delgado from IGN Spain and Joep Crompvoet from KU Leuven for their thorough and constructive reviews of this report.

References

Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. Sci. Am. 2001, 284, pp. 34–43.

European Parliament; Council of European Union. Directive 2007/2/EC Establishing an Infrastructure for Spatial

Information in the European Community (INSPIRE); OJL 108, 25.4.2007; European Parliament & Council of European Union: Brussels, Belgium, 2007; pp. 1–14. Available online: http://data.europa.eu/eli/dir/2007/2/oj (accessed on 20 January 2020)

Kotsev, A., Minghini, M., Cetl, V., Penninga, F., Robbrecht, J. and Lutz, M., INSPIRE - A Public Sector Contribution to the European Green Deal Data Space, EUR 30832 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-41565-7, doi:10.2760/062896, JRC126319.

Masser, I. All shapes and sizes: The first generation of national spatial data infrastructures. Int. J. Geogr. Inf. Sci. 1999, 13, pp67–84

Saligoe-Simmel, J, M Gould, S Sankaran, 2022. Help to make your data FAIR: Part I ArcGIS Blog https://www.esri.com/arcgis-blog/products/arcgis/sharing-collaboration/help-to-make-your-data-fair-part-1/