EuroSDR

European Spatial Data Research

September 2024

Geodata Discoverability (2024)

Joint Workshop of EuroGeographics and EuroSDR

January 16th - 17th 2024, Online

Bénédicte Bucher, Marcin Grudzień, Pekka Latvala, Markus Jobst, Hannes I. Reuter, Hara Papadaki, Jill Saligoe-Simmel, Stine Dau, Raphaëlle Arnaud, Pierlou Ramade, Jordi Escriu, Jeroen Ticheler, Jari Reini

Official Workshop Report

Published by EuroSDR

EUROPEAN SPATIAL DATA RESEARCH

PRESIDENT 2023-25:

Wolfgang Gold, Austria

VICE-PRESIDENT 2023-2025:

Jantien Stoter, Netherlands
Joep Crompvoets, Belgium

SECRETARY – GENERAL 2023 – 2027:

Conor Cahalane, Ireland

DELEGATES BY MEMBER COUNTRY:

Austria: Wolfgang Gold, Gottfried Mandlburger
Belgium: Eric Bayers
Croatia: Saša Cvitković, Željko Bačić
Cyprus: Andreas Hadjiraftis, Georgia Papathoma, Dimitrios Skarlatos
Denmark: Katerina Ritz, Jamal Jokar Arsanjani
Estonia: Tambet Tiits, Artu Ellmann, Evelyn Uuemaa
Finland: Juha Hyyppä, Juha Kareinen
France: Bénédicte Bucher, Aurélien Plyer
Germany: Michael Hovenbitzer, Norbert Haala
Ireland: Guillermo Castro Camba, Audrey Martin
Norway: Jon Arne Trollvik, Ivar Maalen-Johansen
Poland: Anna Bober, Krzysztof Bakuła
Portugal: Mário Caetano, Rui Reis, Marisa Silva
Romania: Iuliana Parvu, Dragoş Badea
Slovenia: Dalibor Radovan, Peter Prešeren, Marjan Čeh
Spain: Julián Delgado Hernández, Jorge Delgado García
Sweden: Tobias Lindholm, Anders Rydén, Heather Reese
Switzerland: André Streilein, Adrien Gressin
The Netherlands: Martijn Rijsdijk, George Vosselman
United Kingdom: David Henderson, Claire Ellul, Jeremy Morley

ASSOCIATE MEMBERS AND THEIR REPRESENTATIVES:

Digitaal Vlaanderen: Jo Van Valckenborgh
Esri: Nick Land, Konrad Wenzel (nFrames)
Het Waterschapshuis: Jeroen Leusink
Hexagon: Simon Musaeus
ICGC: Julià Talaya
IGI: Philipp Grimm
RIEGL: Peter Rieger
Vexcel: Michael Gruber
1Spatial: Dan Warner

CONTACT DETAILS:

EuroSDR Secretariat
Department of Geography
Maynooth University
Maynooth
Co Kildare
Ireland

Tel.: +353 1 7083610
Email: EuroSDR@mu.ie
Web: www.eurosdr.net

B. Bucher, M. Grudzień, P. Latvala, M. Jobst, H. I. Reuter, H. Papadaki, J. Saligoe-Simmel, S. Dau, R. Arnaud, P. Ramade, J. Escriu, J. Ticheler, J. Reini

"Geodata Discoverability (2024)"

Joint Workshop of EuroGeographics and EuroSDR, 16th - 17th January 2024 - Online

Table of contents

GEODATA DISCOVERABILITY (2024)

Joint Workshop of EuroGeographics and EuroSDR

16th - 17th January 2024 - Online

with 11 figures

Bénédicte Bucher [a, b], Marcin Grudzień [c], Pekka Latvala [d], Markus Jobst [e], Hannes I. Reuter [f], Hara Papadaki [g], Jill Saligoe-Simmel [h], Stine Dau [i], Raphaëlle Arnaud [j], Pierlou Ramade [k], Jordi Escriu [l], Jeroen Ticheler [m], Jari Reini [d]

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

[b] University Gustave Eiffel
Champs sur Marne, France

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

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

[e] Bundesamt für Eich- und Vermessungswesen (BEV)
Vienna, Austria

[f] Eurostat
Luxembourg, Luxembourg

[g] Hellenic Cadastre
Athens, Greece

[h] Esri
Indianapolis, US

[i] Agency for Data Supply and Infrastructure (SDFI)
Copenhagen, Denmark

[j] Federal Office of Topography swisstopo
Wabern, Switzerland

[k] Etalab France
Paris, France

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

[m] GeoCat
Ede, the Netherlands

Index of Figures

1 INTRODUCTION

Discovery of geographical data sets is a key step in the process of informing decisions or advancing our understanding of our planet and society based on available data, as well as avoiding the duplication of data. It has led to the development of Spatial Data Infrastructure (SDI) since the late 1990s and to research data portals where research data should be easier to discover and reuse for the sake of open science.

Discovery was initially supported by standard metadata and catalog portals where users express a query by using the very fields of the metadata records and has evolved to adopt paradigms from information retrieval. Advanced functionalities have been added to catalogs that use text fields (title, summary, keywords) of the metadata records to refine or expand query as well as to make recommendations as on data.europa.eu (https://data.europa.eu). In the retrieval paradigm, the user is assisted so that he/she can access a resource that meets his/her need, if that resource exists, even with no prior idea of what the resource looks like. Such assistance is crucial in contexts where users don't have enough knowledge about the domain of the resources to know what discovery solution to use, to express a specific query or to compare the pros and cons of different answers. They are also necessary in contexts where there are so many potential resources that users don't have the time to consider each resource individually. This can be compared to traditional librarians who help the reader who does not know yet which book they would like to borrow. Librarians develop expertise about books and about readers to understand what type of book a user is looking for and to identify specific available books that correspond to that type. They can also indicate to the user more appropriate libraries to go to. Retrieval solutions support query reformulation, refinement, expansion, ranking and clustering of answers on relevance criteria and can also make recommendations. This paradigm has been popularized by Google in the early 2000 to discover websites through a simple search interface. It is applied to different kinds of resources, in particular lately to datasets with for example the dataset search platform of Google.

In that context, a series of online workshops have been co-organised by EuroGeographics and EuroSDR to share experiences and achieve a common vision of what geodata discovery looks like nowadays; this is achieved through the exchange of different perspectives on that field: public or private actors, local or global, Geographic Information specific or not. A first edition called Geodata Discoverability was organized in January 2022 with contributions from users who need to discover data, developers of search engines or of catalog software, developers of open data portals (local to European) and the INSPIRE community (Bucher et al 2023). It evidenced that different users must be considered 1) developers of applications that require data and 2) end users of such applications who will make decisions based on the data. It also evidenced that discoverability of data should be more intertwined with the development and exploitation of the data. As there are many technical solutions, a combination of tools maximizes the chance of potential users to find data relevant to their needs.

A follow up Geodata Discovery workshop[1] was co-organized in January 2024 by EuroGeographics Knowledge Exchange Network on INSPIRE and by EuroSDR Commission on Information Usage, as an on-line workshop. It was called Discovery to insist on a wide scope: how do users identify geodata with potential interest, how do they evaluate their relevance for the application, how do solutions providers help them to do so, in particular how do data providers contribute to such solutions. To extend upon findings of the first edition, we did not consider that discovery ends with downloading or querying data, we consider discovery can continue during the usage when users investigate the potential of the data. A call for proposal was distributed on EuroGeographics and EuroSDR mailing lists as well as towards participants of the previous edition. Emphasis was put on national data providers and how they address the stakes of geodata discovery. Received contributions stemming mainly contributors to discovery solutions and from few users were organized into three sessions.

[1] https://eurosdr.net/workshops/geodata-discoverability-online-workshop
Joint EuroGeographics and EuroSDR Virtual Workshop on Geodata Discovery | EuroGeographics

- First session gathered presentations from a European perspective where discovery is needed to identify potential data sources and their relevance to an application whose scope is bigger than the scope of each individual source. This is the case for example of Eurostat who needs to design panEuropean dashboards based on member state data among other sources.
- Second session gathered presentations from national data providers to assist users in finding their way among the diversity of national data.
- Third session presented more transversal contributions in terms of standards or tools.

The workshop gathered 38 participants from academia, industry or public administration, and from 21 countries (Albania, Austria, Croatia, Cyprus, Denmark, Estonia, Finland, France, Germany, Greece, Ireland, Italy, Montenegro, Norway, Poland, Portugal, Slovakia, South Korea, Spain, Switzerland, United Kingdom).

2 EUROPEAN DISCOVERY SOLUTIONS

2.1 Updates on the creation of pan-European geospatial datasets

Hannes Reuter, Eurostat

GISCO (Geographical Information System of the Commission, https://ec.europa.eu/eurostat/web/gisco) is a permanent service of the European Commission, attached to Eurostat, to provide and integrate geographic data for decision making. More precisely Eurostat needs to systematically discover and reuse different categories data, in particular from authoritative sources, to provide integrated statistics. These categories of data are core geodata provided by member states, new sources of geodata, censuses, surveys, Big Data. At a global level, the United Nation Global Geospatial Information Management group of experts has designed a Global Statistical Geospatial Framework as a way to identify the needed input datasets as well as elements in order to generate support for decisions (UN GGIM 2019). In Europe, the High Value Dataset Regulation enforces member states to release some specific datasets as open data, which are necessary to support these key decisions. Providing detailed specifications of required data is a way to facilitate the discovery process by inciting data providers to adopt a common framework to describe, possibly to structure, their data.

Besides, GISCO reuse authoritative member states datasets to establish pan European datasets which have a range of integration issues, e.g. topology, timeliness. Source data come at different levels of granularity and need to be topologically integrated. GISCO has to establish a relevant workflow for a series of core geospatial datasets. In other words, the discovery of data is tackled by merging data discovery with the identification of relevant tools and methods to process them.

2.2 Enhancing Geospatial Data Discoverability with Ontology and Thesaurus Data in the AquaINFRA Project

Pekka Latvala, Finnish Geospatial Research Institute FGI,
the AquaINFRA project https://aquainfra.eu/

The AquaINFRA project is a European Open Science Cloud (EOSC) project that focuses on developing a virtual environment that contains FAIR data and services that support the research activities in marine and freshwater domains. One part of the AquaINFRA architecture is the Data Discovery and Access Service (DDAS) that provides metadata search and data access mechanisms to various existing data infrastructures.

One component in the DDAS architecture illustrated in (Figure 1) is a python-based web service that aims to improve the discoverability of the services that are connected to the DDAS. The service does

this by enriching the DDAS metadata queries with additional terms that are related to the original query word given by the user. These terms are retrieved by making SPARQL queries to RDF-based ontology or thesaurus data that is hosted on the Apache Jena Fuseki SPARQL server.

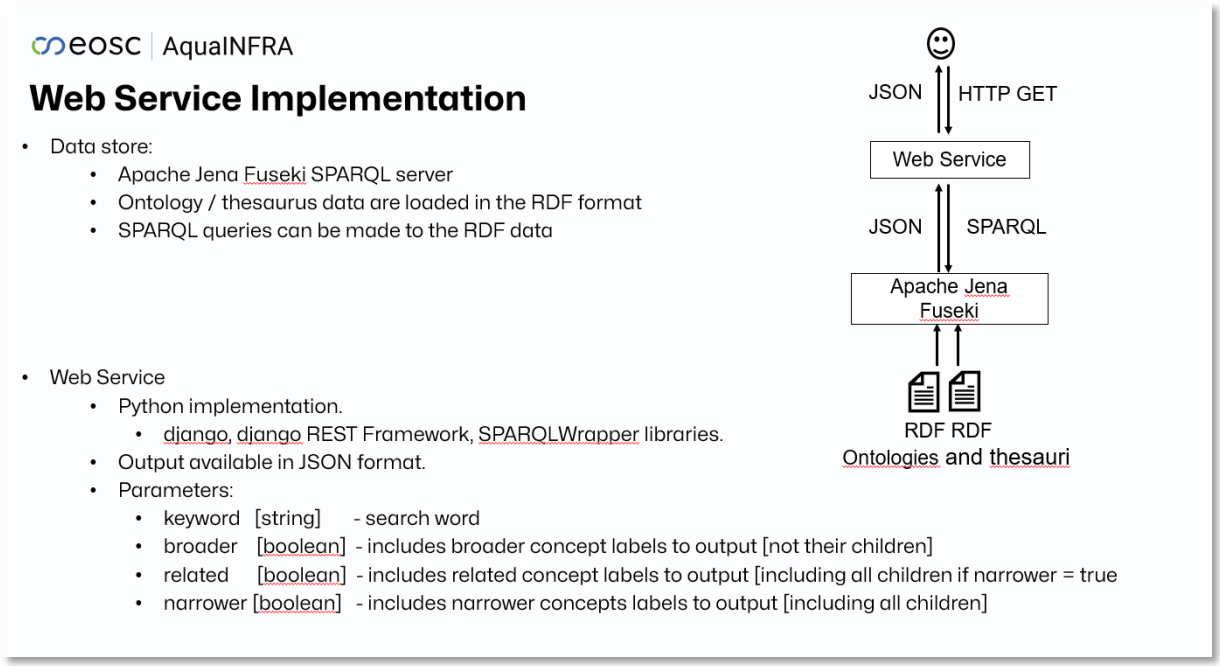

Figure 1: AquaInfra architecture © Latvala

Until now, the work has focused on the GEMET (General Multilingual Environmental Thesaurus) data. The data includes hydrosphere and water themes that are relevant to the AquaINFRA project. The GEMET data contains various concepts that have labels in multiple languages. The web service is focused only on the English language. Each concept may also contain links to other concepts that have "broader", "narrower" or "related" relations.

When the user makes a query to the web service, the server finds first the concepts that have either the skos:prefLabel or the skos:altLabel elements that match with the query word (Figure 2). The web service contains query parameters for including the linked concepts to the output. The broader concepts are limited only to the first immediate parent of the original matching concepts (Figure 3). In general, they may or may not be relevant matches for the query word. Similarly, the related concepts may or may not be relevant matches. The narrower concepts are generally relevant to their broader parent concepts.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT DISTINCT ?s ?label WHERE {
  ?s rdf:type skos:Concept .
  ?s ?p ?label .
  FILTER (regex(str(?label), "lake", "i") || regex(str(?label), "lakes", "i"))
  FILTER (langMatches(lang(?label), "en"))
  FILTER (?p IN (skos:prefLabel, skos:altLabel))
}
```

Figure 2: Example SPARQL query that retrieves all concepts that have a label that contains either the string lake or lakes. © Latvala

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT DISTINCT ?o ?label WHERE {
  <http://www.eionet.europa.eu/gemet/concept/4594> skos:broader ?o .
  {
    ?o skos:prefLabel ?label .
    FILTER (langMatches(lang(?label), "en"))
  }
  UNION {
    ?o skos:altLabel ?label .
    FILTER (langMatches(lang(?label), "en"))
   }
}
```

Figure 3: Example SPARQL query that retrieves the first broader concept of the specified concept. © Latvala

2.3 Enhancing discovery and reuse of national cadastral data through a European Open Cadastral Map

Hara Papadaki, Hellenic Cadastre

Opening data in Europe is expected to yield important benefits to the society and economy, especially in supporting public authorities to carry-out their missions. Member state data that should become open and more reusable are identified as High Value Datasets and are subject to a set of rules in order to ensure free of charge availability. Open geospatial data at a pan-European level aims at improving decision-making at a trans-national scale and supporting directly or indirectly a wide range of applications, such as spatial analysis and planning, environmental monitoring, natural resource management, disaster management and more, across various industries and countries.

In that context, geodata discovery at a European level can be enhanced by the design and implementation of integration workflows to support the derivation of consistent open geospatial panEuropean products based on member state High Value Datasets. This approach is adopted in the Open Maps for Europe

project, led by EuroGeographics. One objective is to provide easy access to open, authoritative, harmonized pan-European cadastral datasets provided by the National Mapping Agencies and Cadastral & Land Registries of Europe (NMCAs).

In particular the project addresses integration of cadastral data from different sources into the Open Cadastral Map (https://www.mapsforeurope.org/datasets/cadastral-all). INSPIRE Directive's specifications and technical guidelines are used to accomplish integration of data coming from different sources. As INSPIRE regulations allow variations and different approaches, some further specifications are necessary for the OCM production process as illustrated on figure 4.

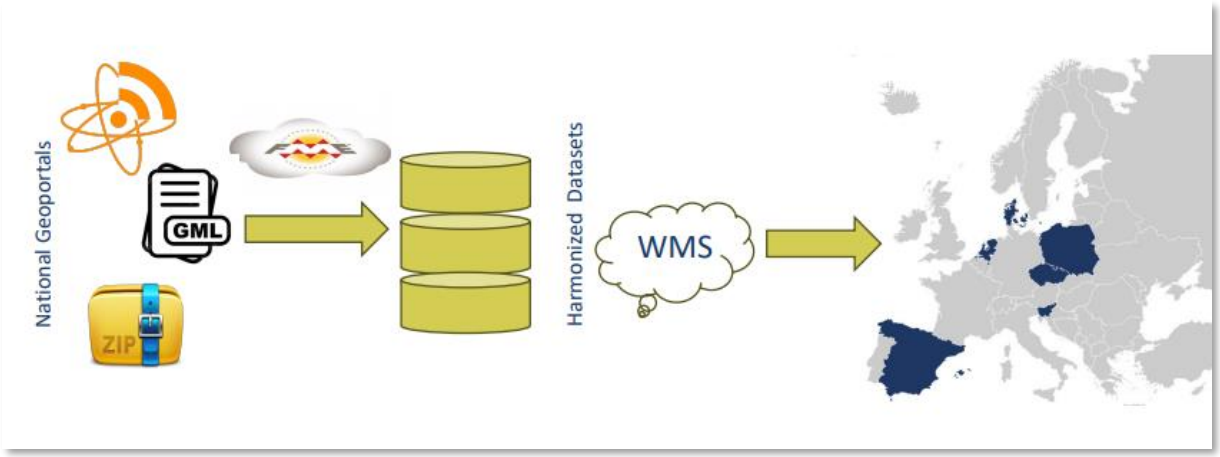

Figure 4: The Open Cadastral Map is fed from each participating member state who need to adapt its INSPIRE services to a more specific and harmonized cadastral format. © Papadaki

For the time being, cadastral map services are available, as a prototype, for six (6) countries, while there is a plan to add nine (9) more countries within the next two years. Those services pertain, depending on the country, information about administrative units, cadastral parcels (and cadastral zones), buildings (and building parts) and addresses. Upon completion of the on-going Open Maps for Europe project, it is envisaged to establish and operate openly a single pan-European map that would depict the principal cartographic features of the countries, such as roads, hydrographic network etc, as well as cadastral maps.

2.4 Towards spatial and open data discoverability for European Data Spaces

Jordi Escriu, Alexander Kotsev (European Commission Joint Research Centre, Directorate on Digital Transformation and Data, Digital Economy Unit)

The latest achievements from European Commission activities with regards location / spatial data discoverability for European Data Spaces include 1) the revamped INSPIRE Geoportal, 2) the support for High-value datasets (HVDs), improving their accessibility and the homogeneity of licencing standards, and 3) aligning INSPIRE and the Open Data communities. Aligning INSPIRE and Open Data Communities is achieved within the extension of the metadata DCAT-AP profile to include HVDs requirements[2], as well as the update of GeoDCAT-AP to define transformation rules from INSPIRE ISO metadata to DCAT3.0 and DCAT-AP for HVDs.

Emerging disruptive technologies that may impact the future of data findability and accessibility. The OGC Rainbow initiative which registers and describes standards and technologies that relate to geographical information will diminish silos currently caused by technologies. An example of a concept

[2] https://semiceu.github.io/DCAT-AP/releases/2.2.0-hvd/

published on OGC Rainbow is “DCAT-datacube”. Generative Artificial Intelligence which is already used to produce code could support the development of questions answering interfaces to discuss the relevance of specific data to meet a given need. The Joint Research Center investigates potential pathways or testing scenarios to issue specific recommendations for INSPIRE to become the Green Deal Data Space in the next years.

3 NATIONAL APPROACHES TO SUPPORT DISCOVERY

Different strategies are developed at the level of national data providers or national open data portals operators to facilitate users discovering relevant data products.

3.1 Discoverability and its vision at Austrian Federal Office of Metrology and Surveying

Markus Jobst, Austrian Federal Office of Metrology and Surveying (BEV)

The core tasks of the Austrian Federal Office of Metrology and Surveying (BEV) in terms of “surveying” are control surveys, maintenance of the cadastre, topographic land surveying and the production of national maps. The resulting products are geospatial base data (or according to the UN-GGIM terminology: fundamental geospatial data themes), which are used in different fields of application, e.g. regional planning, traffic control, environmental analysis, nature conservation, national defense, internal security, agriculture or forestry. For this reason the set of governmental geospatial products is diverse as illustrated in figure 5. One core requirement for the wide usage is its characteristic of stability and reliability, which, to a large extent, defines trustworthiness. Trust is especially important for the collaboration between known and unknown partners, for bridging a variety of use cases and thematic domains or to clarify data sovereignty. The characteristic of trust can easily be supported by appropriate distribution mechanisms, especially if they follow FAIR principles (findable, accessible, interoperable, reusable) and consider modern discoverability methods. These methods exist for the WWW and have been used for decades by search engines. An adoption and embedding for governmental geospatial data is needed to enhance discoverability.

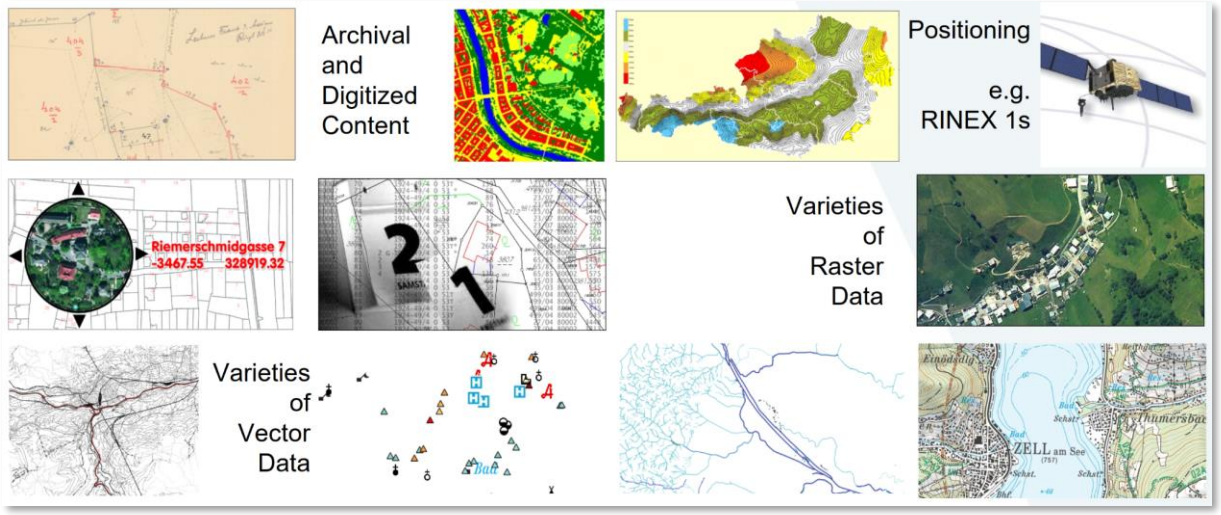

Figure 5: Variety of governmental products.

Accessibility to the geospatial products is a very important step for the process of discovery. Without accessible metadata or products as well as a semantic classification the discovery process can hardly work. Therefore the BEV has a clear strategy and defined principles for data dissemination:

- The publishing date for the products should be consistent with the production process, but also as up-to-date as possible.
- Mass data for the various geospatial data themes are consistent for the whole nation and referenced by a due date.
- All data are in principle free of charge and licensed CC-BY-4.0.
- All standardized IT-services for viewing and dissemination are free of charge.
- All real-time, high value IT-services with complex query and filter options as well as direct integration into customer applications are subject to a fee on the marginal cost principle.

In addition to accessibility, discovery mechanisms require standardized interfaces, well-known formats and most of all metadata at all levels of the IT-infrastructure and geospatial datasets. Metadata have to follow a common consensus about the describing attributes, which are used to describe the dataset as whole, the geospatial features, semantics as well as the IT-service functionality and any other service-oriented architecture method, that is embedded in the dissemination-usage procedure, as illustrated in Figure 6. Open and accessible standards are the fundament for any implementation of a distribution channel.

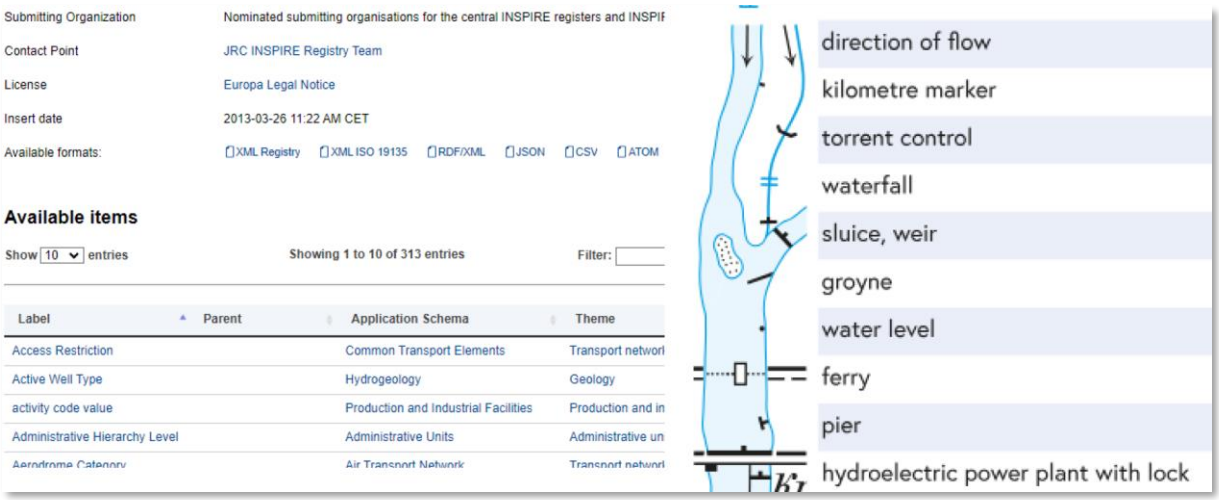

Figure 6: Discovery necessitates not only categories and keywords but also detailed explanation of application schema features, as in a map keys.

Learning from any "Search-Engine-Optimization" (SEO), a controlled, stable and sustainable vocabulary is the main reference for any (geospatial) resource and its features on the WWW. Vocabularies and their ontological relations describe knowledge structures in a formal way. Only this kind of formalization can be used by machines/computers, which therefore understand the geospatial bits and bytes freely accessible on the WWW. From the perspective of BEV more effort has to be focused on a topographical thesaurus and topographic ontological foundations. The aforementioned stability of governmental products, which is also valid for the formal geospatial knowledge representation, could then enhance discoverability and establish trustworthiness.

3.2 Discoverability of Danish Basic Data

Stine Dau, Danish Agency for Data Supply and Infrastructure (SDFI)

The Danish Basic Data program collects and distributes core information about individuals, businesses, buildings, addresses, and geography, so that the public and private sector can use them efficiently. The program continuously aims to enable an easier access to data – users should be able to quickly gain insight into the data (especially its application potential and limitations) and the connection between and the content of the individual registers containing the Basic Data. To this end, two initiatives were brought to life in the Action Plan for Basic Data 2022-24: "Quality improvement of the data model" and "Data declarations of basic data". At the workshop, the approach to and outcome of creating the improved data models and data declarations were presented.

The Basic Data Model was a central element to the initiation of the Basic Data program in 2011. The model still ensures interoperability as it is based on ISO standards and maintained using modeling rules that the data owners themselves have contributed to. Today the model is a useful tool for determining what Basic Data consists of and how the different datasets are related, however, the model is incredibly complex and has only really been targeting experienced and technically advanced users. It has proven insufficient in reaching potential new users who wish to discover the data at a more basic level.

To make Basic Data more discoverable, the models have undergone a significant upgrade in quality as additional rules for conceptual modeling have been embedded in the Danish Basic Data program in the recent two years. As a result, the upgraded models can, among other things, be displayed as object catalogs as illustrated in Figure 7. These catalogs provide an overview of the object types, including the properties of and relationships between the different basic data types. In addition, the object catalogs show relevant definitions, descriptions, sources, diagrams and information on how/where data can be retrieved from the data distributor. These definitions concern concepts represented in the data like wetlands, forests, buildings. Additionally, users can now search for terms and definitions, and thereby discover the relevant data in a comprehensive Basic Data dictionary. Users can also discover relevant datasets using an overview of the business concepts in Basic Data and their context, which also provides access to the new object catalogs.

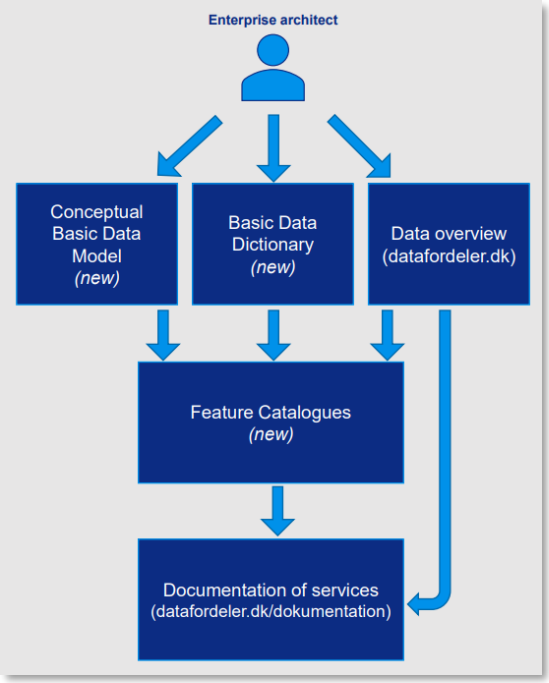

Figure 7: Different journeys to discover data.
© Agency for Data Supply and Infrastructure

The purpose of the "Basic Data Declarations" on the other hand is to provide even more general insight into Basic Data, with focus on its data quality, application potential and limitations. The aim of the declarations is to increase the understanding of the existing data and thus enhancing the potential for creating value with Basic Data. The declarations also seek to contribute to creating visibility and awareness about the data quality of Basic Data.

New and potential users are provided with a quick overview of the registers, their content and any limitations for use. At the same time, the declarations compile a list of references to more detailed or specific documentation, such as the object catalog, the specification or reports on data quality, which are useful for more experienced users. The outline and content of the data declarations were developed meticulously in cooperation with existing and potential users and have received various positive responses from both data authorities and users.

3.3 Discoverability on the French national open data portal

Pierlou Ramade, Etalab, France

Data.gouv.fr (https://www.data.gouv.fr) is the platform where all French administrations have to publish their public open data. The description of data schemas is a useful item to enhance discoverability and interoperability but is often lacking. In that context, Etalab which operates data.gouv.fr platform has set up an ecosystem of services that fosters the documentation of schemas and the reuse of existing schemas for the publication of open data.

This ecosystem is composed of:

- the catalog of public schemas (https://schema.data.gouv.fr/) which is enriched whenever a new schema emerges, after rounds of concertation and drafting (all schemas are hosted on Github), a schema is more precisely a json file that describes the structure of other files published as open data.
- a platform to assist data producers in publishing data according to a selected schema as many producers are not data experts (https://publier.etalab.studio),
- a platform to verify whether a file conforms to a selected schema (https://validata.fr/).

There is for example a public schema for electric vehicle charging stations. To get a subsidy, entities that have charging stations have to declare them on data.gouv.fr, following this dedicated schema. The link to the schema is in the law. More than 1100 files are consolidated into a single file recreated everyday with potential newly uploaded ones.

Another initiative of data.gouv.fr to facilitate discovery and reuse is the publication of data reusages. Usages of data can be the production of a map for located data or the derivation of specific aggregates. They are associated with tags, like for example "education", "economics", "politics", that can be used to discover them through the portal explore.data.gouv.fr (https://explore.data.gouv.fr). When selecting a specific usage, one can read a detailed description of the motivation for this usage and the methodology and source data.

3.4 Geodata discovery strategy in Switzerland: the geocat.ch portal

Raphaëlle Arnaud, Federal Office of Topography swisstopo, Switzerland

In collaboration with all stakeholders (Confederation, cantons, municipalities, etc.), the "Swiss Geoinformation Strategy" aims to make reliable, detailed, up-to-date and interoperable geoinformation available to all users in a simple, interconnected and, where appropriate, real-time manner. Accessibility includes discoverability.

Metadata is essential for making data more accessible, and must therefore be standardized, of high quality and easy to use. Switzerland was the first country in Europe to introduce in federal law the obligation to document geodata in the form of standardized metadata with the Geoinformation Act that entered into force in July 2008. This objective is still part of the Swiss geoinformation strategy, and currently includes more specifically the revision of the Swiss metadata standard (future eCH-0271) and the improvement of the search interface of the Swiss geometadata catalog - geocat.ch (https://www.geocat.ch) illustrated in figure 8.

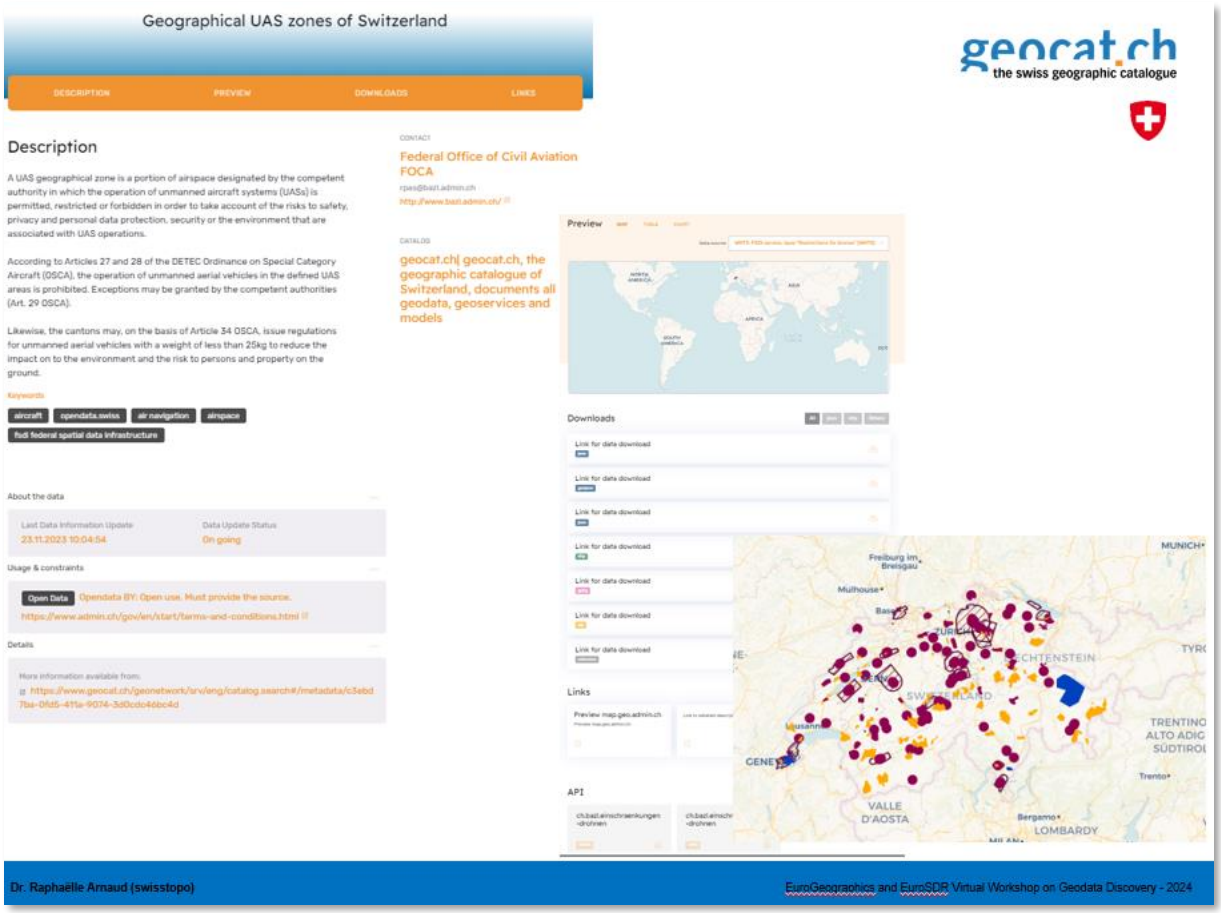

Figure 8: Simplified User Interface on the Swiss geographic catalog.

The user can express his query as a keyword, like for example “airspace zone”, and the system retrieves all records whose metadata contains these keywords, in the title, the description, or the keywords. When the user selects a record, like for example “Geographical UAS zones of Switzerland” he has access to rich metadata to discover the data content.

4.1 Discovery of geospatial resources by OGC API Records with collections of descriptive information about the resources (metadata)

Jari Reini, National Land Survey of Finland

OGC API Records is a multi-part specification developed by the Open Geospatial Consortium (OGC) to offer the capability to create, modify and query metadata on the Web. It provides standardized access and query of metadata records for geospatial data and services, as illustrated in figure 9, and richer metadata about geospatial resources, including information about data quality, spatial extent, temporal coverage, and more.

Path Template	Relation	Resource
Common		
{root}/	none	Landing page
{root}/api	`service-desc` or `service-doc`	API Description (optional)
{root}/conformance	`conformance`	Conformance Classes
{root}/collections	`data`	Metadata describing the spatial collections available from this API.
{root}/collections/{collectionId}		Metadata describing the collection which has the unique identifier `{collectionId}`
Records		
{root}/collections/{collectionId}/items	`items`	Search results based on querying the service for records satisfying 0..n query parameters.
{root}/collections/{collectionId}/items/{recordId}	`item`	Record of metadata which has the unique identifier `{recordId}`.

Figure 9: OGC API Records endpoint

This specific OGC standard is a promising cornerstone to enhance the sharing and reusing of metadata and more generally the development of solutions that use metadata, such as discovery solutions.

4.2 Ensuring successful geodata discovery and reuse through geospatial product thinking

Jill Saligoe-Simmel, Esri

By adopting a geospatial product thinking mindset, organizations that provide data can design data products that meet the needs of data consumers, leading to successful discovery and reuse as illustrated in Figure 10. To do so, data providers must understand the diverse needs and expectations of data reusers, including easy discoverability, seamless integration into workflows, and other specific requirements, to tailor products.

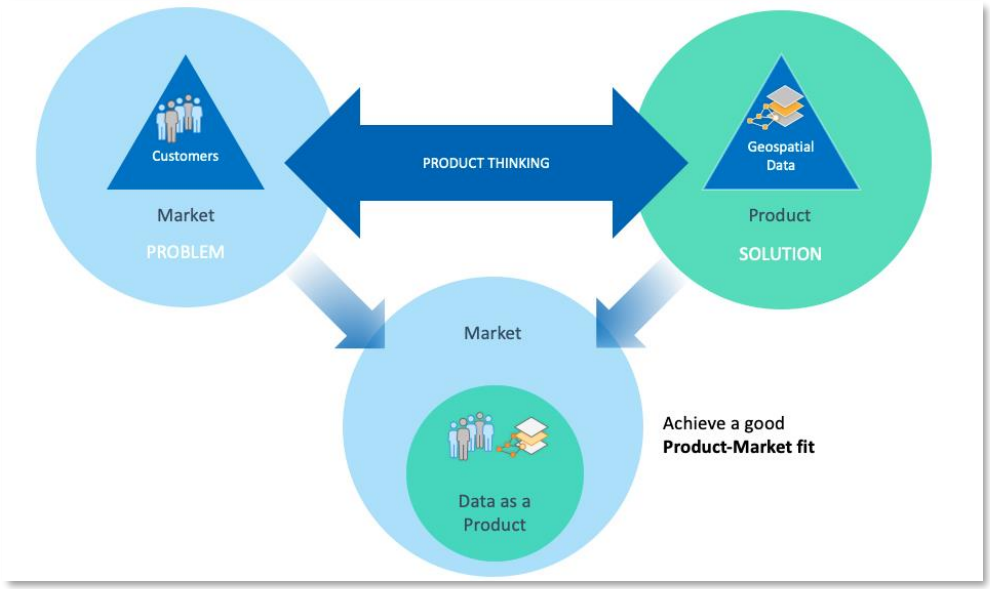

Figure 10: Product thinking is a journey between the problem space of the users and the solution space of your business. The goal is to reduce the gap and achieve a good product-market fit. Graphics inspired by (Katakam 2020)

Adopting a product thinking approach involves user-centered design and continuous improvement based on user feedback. Data products should be easy to discover, understand (configured for quick visualisation), and use (i.e. self-describing, interoperable, licensed for reuse). They should be relevant (analysis ready, with key fields for joining with other data), reliable (incl. persistent), and accessible in various forms to serve a broad audience, including non-technical users. By making data ready for friendly reuse, data providers can empower data reusers to seamlessly integrate the data into their workflows and derive meaningful insights.

With modern Web GIS patterns, the trend worldwide is an expanding customer base as illustrated in Figure 11. Consider, for example, the expanding human audiences of Architect, Engineering, and Construction (AEC) professionals discovering and reusing geospatial data directly from within their design software. Consider also the rapidly emerging role of machines as geodata consumers in sensor networks, smart cities, and as natural language AI assistants.

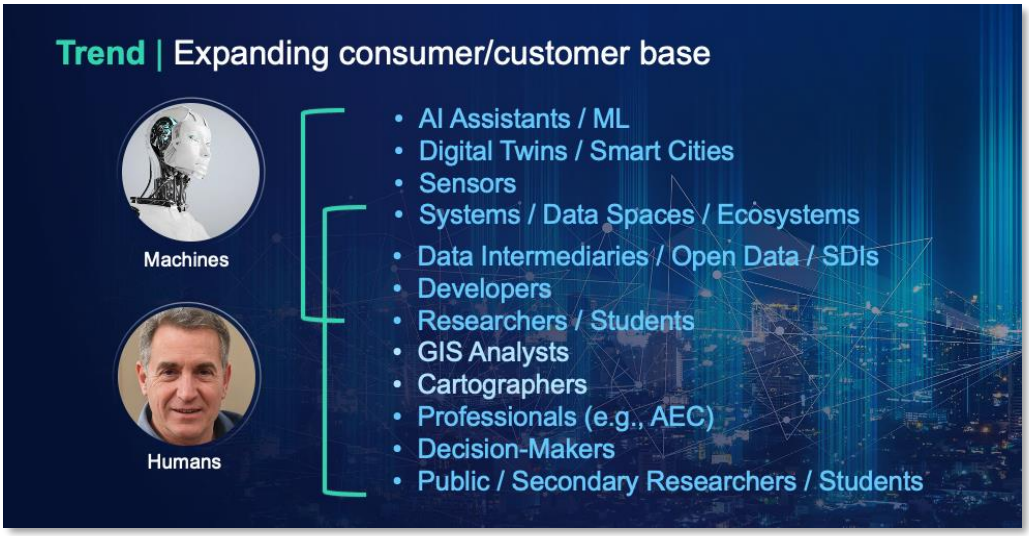

Figure 11: ESRI vision of human and machines users.

The EU Open Data Directive presents an exciting opportunity for data providers to position themselves as leaders in the market. By embracing this directive, data providers can contribute to the positive impact of open data. The directive emphasizes focusing on outcomes, delivering exceptional value, and enabling data reusers to make informed decisions.

Web GIS and modern geospatial infrastructure support the goals of the EU Open Data Directive. For example, today's geospatial data hubs are open data content-sharing, collaboration, and stakeholder engagement systems that maximize data discoverability and reuse. By leveraging out-of-the-box tools and applying pragmatic good practices for authoritative data providers [ESRI 2023], organizations can quickly advance the discoverability and reuse of their data products, ultimately empowering consumers to innovate and drive positive change.

5 CONCLUSION AND PERSPECTIVES

Geodata discovery still is a complex challenge, from a user perspective but also for all actors currently involved in that field. It is challenging to know what geographical data exist, and for a start what kind of geographical data can be found and where to find them. It also is difficult to assess, in a comparable way, the relevance of different data for a specific usage. This has to do with the notion of geographical data quality.

Different actors contribute to the field of geodata discovery. These can be data providers who wish their data to be more visible and used. Search engines or data portals wish their search solution to be more used by providers and by users. Standardization bodies advance transversal technologies, mainly metadata, to be adopted by both groups.

The discovery process can be improved by a better understanding of users and market and adaptation to users. This adaptation is now addressed not only by portals or engines but also by data providers who design user oriented geospatial products based on their available data, like for example a panEuropean cadastral product. It can also be achieved by designing users' journeys through graphical interfaces to assist users in exploring the complexity of geodata products, including quality aspects. These journeys can be organized within a national portal or across data spaces.

Many metadata can contribute to a successful discovery process and must be better managed and exchanged. For example schema is an important metadata item to enhance discoverability as it contains more semantics than the name of the dataset or datasource, but its documentation is often lacking.

Users currently often have to identify search portals priori to search for data. It is necessary to make metadata more interoperable and exchangeable, thanks to standards like the OGC API records, so that search engines and catalogs can forward user queries to one another.

Last, there is a knowledge gap between the capacities of advanced tools like Geonetwork to exploit metadata and their actual usage.

Acknowledgements

The authors wish to thank EuroGeographics and EuroSDR for their support in organizing the workshop as well as more specifically the following persons. Matina Fuentes, Patricia Sokacova, Giuseppe Novella and Alina Tapova from Eurographics ran the online workshop and organized the publication of presentations in a smooth and efficient way ensuring people participation. Claire Ellul, from University College London, and Frédéric Cantat, from the French national forest and geographic information institute (IGN France), made a thorough and constructive review of a first version of this report.

References

Bucher B., Grudzien M., Delattre N., Escriu Paradell J., Folmer E., et al. (2023): Geodata Discoverability. EuroSDR Office of Publication. ws_report_geodatadiscoverability.pdf (eurosdr.net)

Katakam N. (2020): Product Thinking 101. Medium article, https://uxplanet.org/product-thinking-101-1d71a0784f60

Esri. 2023. Best Practices for Authoritative Data Providers: Maximize the deliciousness of your shared geospatial content. https://storymaps.arcgis.com/stories/073f2420e5b24a06968973477bda2914

UN GGIM, 2019: The Global Statistical Geospatial Framework, New York
https://www.efgs.info/wp-content/uploads/2019/10/The_GSGF_NY_2019.pdf