EuroSDR

European Spatial Data Research

May 2024

EuroSDR Knowledge Graph about Geodata Products in Europe

EuroSDR Linked Data Project Mid-Term Report

Bénédicte Bucher, Pekka Latvala, Eva Bookjans, Nicolas Bus, Audrey Bouet, Erwin Folmer, Marjan Ceh, David Portoles, Luis Martinez Lorenzo, Dimitris Kotzinos, Hara Papadaki, Pasquale Di Donato

Official Publication Nº 76

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: to be confirmed
Cyprus: Andreas Hadjiraftis, Georgia Papathoma, Dimitrios Skarlatos
Denmark: Katerina Ritz
Estonia: Tambet Tiits, Artu Ellmann, Evelyn Uuemaa
Finland: Juha Hyyppä, Juha Kareinen
France: Bénédicte Bucher, Aurélien Plyer
Germany: Michael Hovenbitzer, Norbert Haala
Greece: Lolonis Panagiotis
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, Adrien Gressin
The Netherlands: Martijn Rijsdijk, George Vosselman
United Kingdom: David Henderson, Claire Ellul

ASSOCIATE MEMBERS AND THEIR REPRESENTATIVES:

Digitaal Vlaanderen: Jo Van Valckenborgh
Esri: Nick Land, Konrad Wenzel (nFrames)
Field Group: Leif Erik Blankenberg
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

The official publications of EuroSDR are peer-reviewed

Bucher B., Latvala P., Bookjans E., Bus N., Bouet A., Folmer E., Ceh M., Portoles D.,
Martinez Lorenzo L., Kotzinos D., Papadaki H., Di Donato P.

"EuroSDR Knowledge Graph about Geodata Products in Europe"

EuroSDR Linked Data Project Mid-Term Report

Table of contents

EUROSDR KNOWLEDGE GRAPH ABOUT GEODATA PRODUCTS IN EUROPE

EuroSDR Linked Data Project Mid-Term Report

with 11 figures

Bénédicte Bucher [a *, b], Pekka Latvala [c], Eva Bookjans [a], Nicolas Bus [d], Audrey Bouet [d], Erwin Folmer [e, f], Marjan Ceh [g], David Portoles [h], Luis Martinez Lorenzo [h], Dimitris Kotzinos [i], Hara Papadaki [j], Pasquale Di Donato [k]

[a] National Institute for Geographic and Forest Information –IGN
Saint Mandé, France
[*] benedicte.bucher@ign.fr

[b] University Gustave Eiffel
Champs sur Marne, France

[c] National Land Survey
Masala, Finland

[d] Centre Scientifique et Technique du Bâtiment
Sophia Antipolis, France

[e] Kadaster
Enschede, The Netherlands

[f] University of Twente
Enschede, The Netherlands

[g] University of Ljubljana
Ljubljana, Slovenia

[h] Idearium Consultores
Zaragoza, Spain

[i] CY Cergy Paris University
Cergy Montoise, France

[j] Hellenic Cadastre
Athens, Greece

[k] Federal Office of Topography swisstopo
Wabern, Switzerland

Index of Figures

1 CONTEXT

1.1 Sharing and reusing geospatial data: a continued challenge

Spatial and geographical data as well as spatial referencing frameworks are key assets for the green transformation of our societies. They correspond to a diversity of scopes and technologies (Masser and Crompvoet, 2018). These can be topographical data, digital elevations models, land cover or land use maps as well as statistical surveys using spatial grids. The digital transformation of society and the trend towards open data have exponentially increased the amount of data publicly available, for example Web content, digitised archives, social media, many of which have some spatial characteristics. All these data help us to visualise and better understand the world around us, for example monitor crops or evaluate urban sprawl. They are important tools in the decision-making process for both private and public sectors. Combined this data can become even more valuable than it already is, for example to detect urban heat islands and to evaluate vulnerable population living there (Masson et al., 2020). Space often acts as a common referencing framework to integrate different information into a spatial analysis model, which can show spatial correlations and predict future trends. Such common referencing framework can be direct – if the information can be associated with geodetical coordinates – or indirect – if the information can be associated with entities which themselves have coordinates. Examples of entities that support indirect spatial referencing are administrative units, addresses or place names.

There is a need for more assistance in discovering and exploring available data as well as integration technologies, and eventually selecting data and integration technologies to prepare data that can support the decisions at stake (Bucher et al., 2023) (Zrhal et al., 2021). This discovery and reuse process require users to be familiar with different search portals and different integration and analysis technologies. They even may need to download the data to better assess if potential uncertainties propagate in their application and affect the result. Altogether, this can be so challenging that users may consider producing new data instead, or simply using the data easiest to find even if it isn't the most relevant and reliable. Conversely, data providers are faced with the challenge of making their data visible to potential users and receiving feedback on the pertinence and usability of their data.

These challenges are particularly experienced by national mapping and cadastral agencies (NMCAs). They lack efficient solutions to share data internally within their organizations in a context where multiple data are derived from core products. They also need to discover and reuse external data sources relevant to their activities, e.g. to get to know, in between aerial campaigns, what changes occur that should be propagated in core products, like the development of a wasteland. They need to monitor more systematically the effective usages of the data they publish, which could be useful to report on the value of the product and to improve data to better meet users' needs.

1.2 Semantic Web meets Spatial Data Infrastructures

Two main domains of research, innovation and operation have addressed these issues, namely: 1) the domain of Spatial Data Infrastructures (SDI) and 2) the Semantic Web, which includes the Web of Data. In this section, we recall briefly SDI's contributions and describe more in detail the Semantic Web, incl. the Web of Data, as readers may be less familiar with these domains.

Since the late nineties, spatial data sharing and reuse has been addressed by the Geographic Information communities with the design of SDI and the establishment of interoperability standards for key geoservices such as geodata cataloging, geodata viewing, geodata downloading, and geodata processing, largely promoted by the Open Geospatial Consortium (OGC) community. Specific metadata models have also been defined to handle the specificities of geographical data: the ISO 191xxx series which includes ISO19115 for geographical metadata and ISO19157 for quality

metadata. In Europe, the European Union Directive 2007/2/EC (INSPIRE) requested all member states to contribute to a European SDI to support the reuse of member state data to design and monitor European environmental policies. It has yielded foundations to share and reuse member states environmental data across the European Union, like standards profiles, data services, experts' networks and governance. These foundations are important assets to implement the European New Green deal and the digital transformation of Europe.

At the same time when SDI emerged, the Web shifted from a situation where content is published and accessed as documents towards a new vision where content is also shared as data, sometimes designated as "raw data" as opposed to Web pages built from the raw data. This new Web of Data is meant to support the reuse of Web content but also of any data. This was a timely evolution to support the open data and the open science trends which adopted the Web technologies to publish and share datasets. (Wilkinson et al., 2016) coined the expression of FAIR principles for Findability, Accessibility, Interoperability and Reusability. The challenge of a Web of Data is detailed by (Heath and Bizer, 2011) as follows:

- "How best to provide access to data so it can be most easily reused?
- How to enable the discovery of relevant data within the multitude of available data sets?
- How to enable applications to integrate data from large numbers of formerly unknown data sources?"

The technical approach to achieve a Web of data is established by the World Wide Web Consortium (W3C). The W3C identifies best practices for content providers: posting online content in the form of raw data instead of web pages, using open formats, structuring the data, and linking with other content.

A W3C standard has been proposed to describe datasets on the Web: the Data Catalog Vocabulary (DCAT). This standard is generic and has been largely adopted by open data projects. Besides,

Fortunately, Semantic Web and SDI communities have collaborated to define joint standards. Over the last 5 years, the joint W3C/OGC Spatial Data on the Web working group has been studying how to tackle the specificities of geographical information on the Web. Its findings are published in a rolling document (Archer et al., 2023) containing many recommendations applicable to the production and delivery of linked spatial data and metadata in a web environment.

Additionally, the Semantic Web focuses on improving the interaction with users:

- Considering their queries as concepts, "things not strings", to improve the relevance of answers.
- Dialoguing with users to better adapt to their context.

These interaction capacities are supported by Knowledge Graphs (KG). A KG is a "graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent potentially different relations between these entities. The graph of data (a.k.a. data graph) conforms to a graph-based data model, which may be a directed edge labelled graph, a heterogeneous graph, a property graph, and so on" (Hogan et al., 2021). KGs adopt graph models to be closer to the mental model. Formally, such models are associated with "description logic" reasoning capacities. That allows for a more precise automatic 'classification', for example, the classification of a user into a more fitting profile group, so that a more relevant recommendation can be given to that user. The GAFA industries were pioneers in developing their own KGs to improve their capacity to attract users and to generate more activity on their respective web platforms. Open Knowledge Graphs are being developed, the most famous being Wikidata, which is a collaborative Knowledge Graph used as a core structure by the Wikimedia foundation projects and DBpedia, which extract structured content from the information available through Wikipedia. The Semantic Web continuously evolves at the junction between the open KG trend, the Web of Data trend as well as information retrieval. Core principles are promoted either to contribute to an open KG, or to structure data or metadata, or to annotate Web pages for search engine optimisation:

1. The adoption of http uniform resource identifiers (URIs) referencing things of interest -e.g. an organization, a datatype, an entity-,
2. The creation of human-readable and machine-readable descriptions for these things of interest
3. Linking the URIs to these descriptions makes them readily available
4. Applying 1), 2) and 3) to the things used to build the description – e.g. a datatype, an attribute, a controlled vocabulary-

1.3 EuroSDR past Semantic Web activities

Since 2014, EuroSDR has been investigating possible synergies between the Semantic Web (SW) and National Mapping and Cadastral Agencies (NMCAs) with two distinct motivations:

- Can SW technologies, like Linked Data, perform better than the current techniques used by NMCAs for issues such as multiple representation, schema transformation, and cataloguing? If so, how can NMCAs adopt these techniques?
- Does the SW require a framework for spatial data? If so, how can NMCAs, with their experience in handling and sharing spatial data and its particularities, contribute to such a framework?

EuroSDR SW activities were organized around working seminars, with a limited number of participants, and bigger open events listed below:

- 2019: SDI.next Linked Spatial Data in Europe [1],
- 2020: Knowledge Graph in Action, DBPedia, Linked Geo-Data and data integration, co-organized with DBPedia and joint with EuroSDR Volunteer Geographic Information activity[2],
- 2023: Joint EuroSDR-PLDN-GeoE3 Symposium "Metadata Matters, What Role Does Linked (Meta)Data Play in the Geospatial World?"[3] .

During these exchanges, obstacles to the dialogue between the Web of Data community and NMCAs were identified (Bucher et al., 2020). NMCAs, on the one hand, are officially mandated to produce reliable geodata and reduce any potential uncertainty in user decisions relying on geodata. On the other hand, Web of Data developers prioritises data discovery and reuse in new applications and the reduction of any possible obstacles to data download and processing. A conflict of interest arises if application developers improperly use data, even if it is done unknowingly. Another conflict of interest arises from having different authority frameworks: legal mandate and obligations vs the number of users, quality of data vs quantity of data.

An EuroSDR Linked Data Technical Group was created to maintain a more continuous dialogue between experts from the Semantic Web domain and the Geospatial Information domain, including a few experts familiar with both domains. The group defined shared procedures for transforming available ISO 19139 metadata into RDF files, loading these files into a common sandbox, and creating

[1] SDI Next: Linked Spatial Data in Europe | EuroSDR (https://www.eurosdr.net/workshops/sdi-next-linked-spatial-data-europe)
[2] https://www.eurosdr.net/workshops/knowledge-graph-action-dbpedia-linked-geodata-and-geo-information-integration
[3] https://www.eurosdr.net/workshops/joint-eurosdr-geoe3-pldn-symposium-metadata-matters-what-role-does-linked-metadata-play

links between the records based on keywords and topic categories to achieve a draft Knowledge Graph of Digital Assets in Europe.

At the end of 2022, a project was launched to prototype a more advanced EuroSDR Linked Data proof-of-concept meeting the following requirements:

1. The prototype should cover the knowledge life cycle, i.e. the production of nodes, links, and queries, to better identify potential benefits as well as costs.
2. Use cases should efficiently improve the work of the participants and be based on real needs.

The main objective of this report is to present the focus of the prototype, in particular the specific use cases that have been defined within the participants, to the EuroSDR community. It also aims at explaining the relevance of Semantic Web, Linked Data and Knowledge Graph to readers who are familiar with the stakes of publishing and sharing geographical data, all particularly mapping and cadastral agencies, but who don't have prior expertise in these Semantic Web technologies.

The remainder of this report is structured as follows: Section 2 provides an overview of recent developments in the adoption of Semantic Web paradigms by the EuroSDR community. Section 3 describes the target KG.

2 UPDATES ON THE ADOPTION OF SEMANTIC WEB PARADIGMS FOR GEOMETADATA AND GEODATA

This section presents the updates on the adoption of SW paradigms reported during the Linked Data group meeting and the 2023 Metadata Matters Symposium.

2.1 RDF Metadata for cross domain catalogs and mediators

RDF metadata is increasingly applied in catalogs. Catalogs or search engines allow users to discover data from different providers thanks to a common metadata API. The RDF paradigm is used to create links between metadata and design cross-domain catalogs. It is of particular interest to organizations concerned with archival tasks as in the network of terms approach developed at the Dutch National Digital Heritage program (NDE) to develop an infrastructure to facilitate the discovery and reuse of archives across different domains, see Figure 1 and Figure 2.

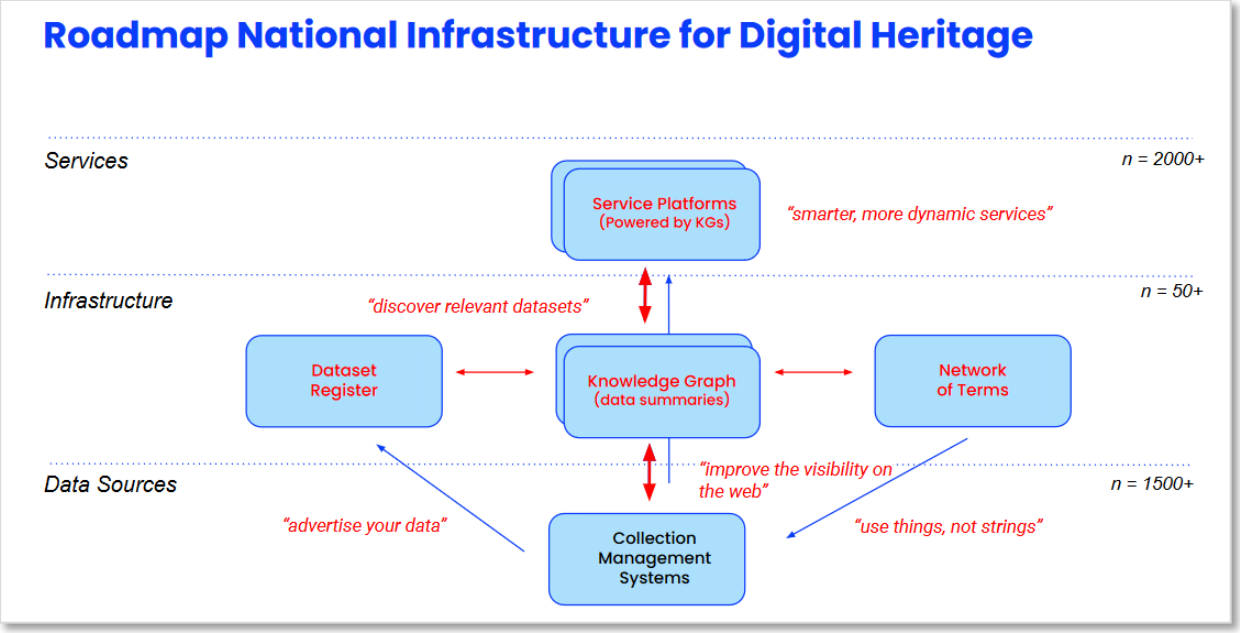

Figure 1: The Dutch National Digital Heritage (NDE) infrastructure associates Knowledge Graph to registries to develop cross domain discovery and reusage of archives (© Meijers[4]).

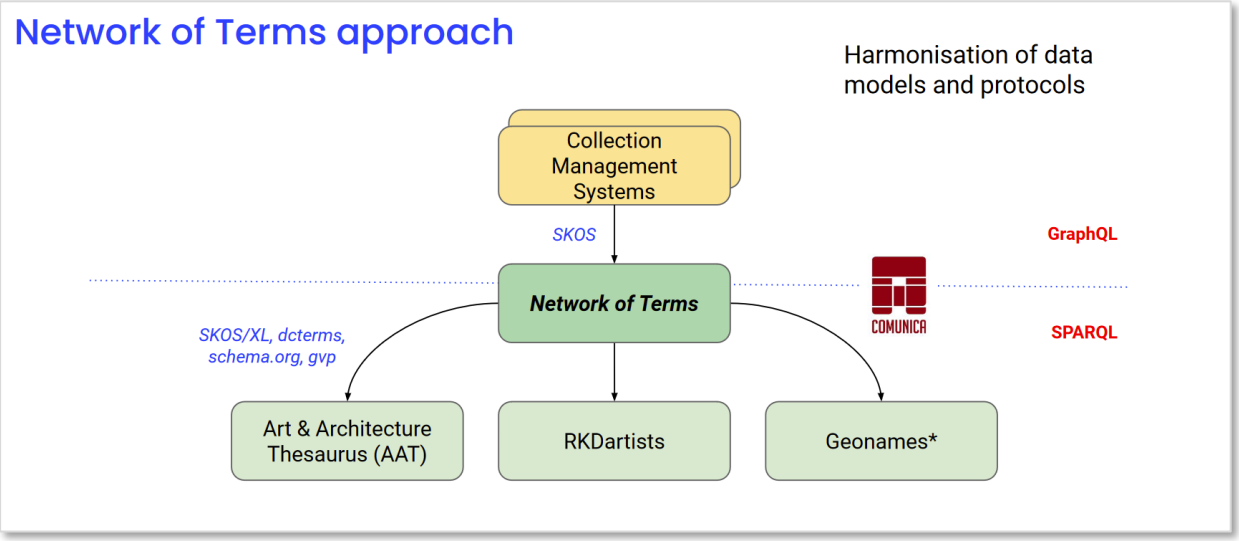

Figure 2: NDE Network of Terms.
A key step to develop cross-domain access to resources in the NDE program is the harmonization of metadata achieved through the "Network of Terms" graph that links terms across resource providers thanks to semantic Web technologies in the NDE program (© Meijers[5]).

4 https://www.pldn.nl/wiki/Metadata_Matters,_What_Role_Does_Linked_(Meta)Data_Play_in_the_Geospatial_World%3F_%E2%80%93_6_September_2023,_Sound_and_Vision,_Hilversum,_the_Netherlands
https://www.pldn.nl/images/pilod/0/07/2_TermenNetwerk_EnnoMeijers.pdf page 12

5 https://www.pldn.nl/images/pilod/0/07/2_TermenNetwerk_EnnoMeijers.pdf page 17

Integrated catalogs can also focus on a given domain of interest observed through different data technologies. The AquaINFRA project is a European Open Science Cloud (EOSC) project that aims to develop a virtual environment containing FAIR data and services supporting marine and freshwater research. The project is developing an ontology-based search component that is a web service providing related terms for a given query word. The related terms are retrieved via SPARQL queries from the RDF-based ontology or thesaurus data. The found terms can be added to the metadata queries for better search results.

RDF metadata and KG are also used in mediators to allow users to query several sources with a single language or interface. The mediator automatically translates the user query expressed in this single federated model into queries adapted to each source. Such solutions can also rely on RDF models to consolidate the answers received from the different sources and send a unique consolidated answer to the users. As part of the GeoE3 project, this has been tested for cross-border applications, as well as for real estate applications, which gather as much data as possible on the buildings of a given parcel from a variety of different sources using the IMX architecture (see Figure 3 below and https://github.com/Geonovum/WaU).

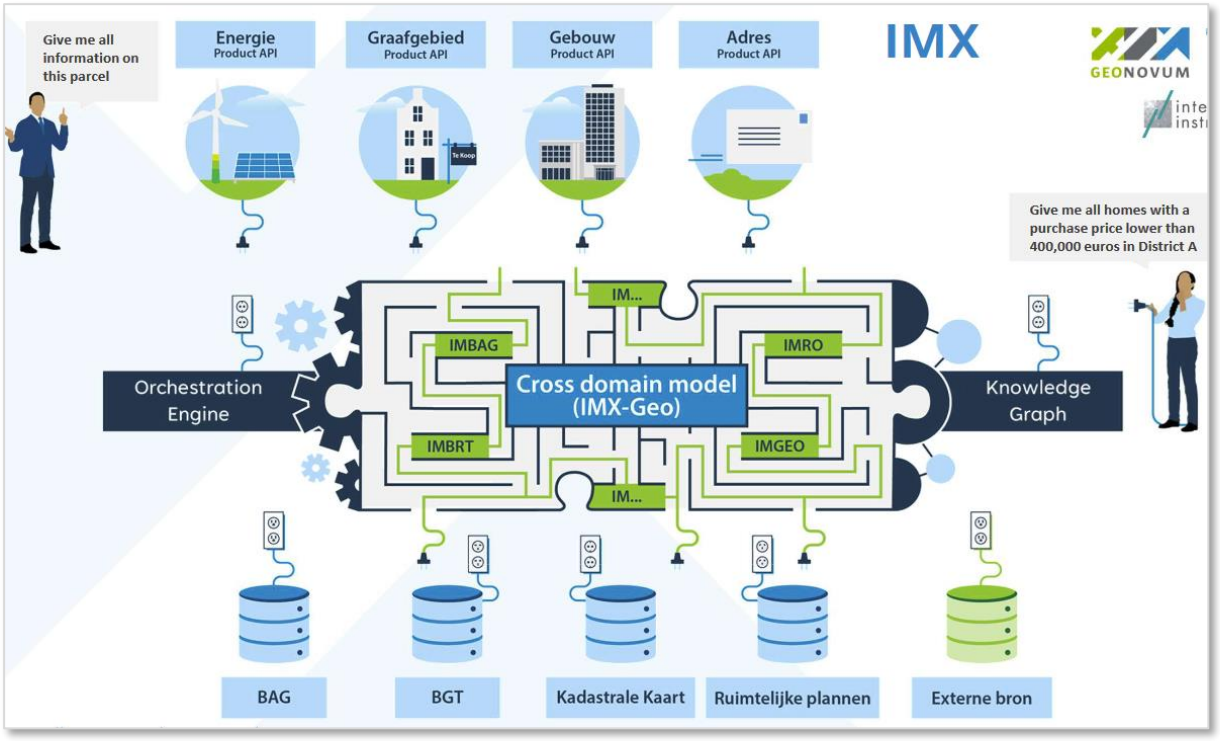

Figure 3: The IMX architecture developed by Dutch Kadaster.
A data brokering engine relies on a KG to interpret a user query into queries on different sources (©Kadaster[6]).

6
https://www.pldn.nl/wiki/Metadata_Matters,_What_Role_Does_Linked_(Meta)Data_Play_in_the_Geospatial_World%3F_%E2%80%93_6_September_2023,_Sound_and_Vision,_Hilversum,_the_Netherlands
https://www.pldn.nl/images/pilod/2/20/4_IMXGeo_PanoMaria.pdf page 6

Metadata and ontologies are also used to improve the dialoguing capacities or catalogs or even GIS through question-answering solutions. Simon Scheider from Utrecht University develops a proof of concept where users ask questions to a system that answers through a map generated from different data sources and GIS workflows. Another example presented at the Metadata Matter Symposium is a system using Q&A to query underground and land use data to predict excavation damage.

Datasets harmonisation, i.e. the modification of datasets so that they can be more easily manipulated and handled together, also benefits from RDF metadata. This typically involves transforming schemas to diminish syntactic heterogeneities or changing geometries to ensure features overlay correct. On open data platforms, the use of existing schemas to structure data before its creation or publication, as well as the publication of the utilised schema for reuse, is encouraged. The schema.org platform is often mentioned as a core platform to achieve such harmonization.

Another approach developed on the Swiss open data portal involves asking open data providers to fill in detailed linked metadata that can be processed to automatically generate datacubes which supports on-line analytical questions.

2.2 RDF Data supporting integration

RDF is also adopted at the level of data to support integration.

Dataset interconnection can be the creation of matching links between homologous features from two different datasets, for example links between data describing a specific building in an authoritative topographic data product, data describing the same building in DBpedia and data describing the same building in OpenStreetMap. Dataset interconnection can also be the creation of links between features, that correspond to relations between these features, for example a building that has a postal address and is within a parcel. Sets of links, sometimes called linkset, are similar to a table joining service. They can be published as RDF data even if the original interconnected data are not RDF data themselves, provided there exist external keys, preferably URIs, for the interconnected features in each dataset.

Geographical knowledge graphs develop as stand-alone RDF datasets that contain links between features but also main properties for these features. Different examples of geographical KGs can be found in the EuroSDR official publication 73 (Bucher et al., 2021). The Kadaster geographic knowledge graph is used to improve user experience when accessing public information transformed into interconnected RDF data (Ronzhin et al., 2019). It is built by interconnecting the three open base registers from Kadaster as well as datasets from outside Kadaster as illustrated in Figure 4.

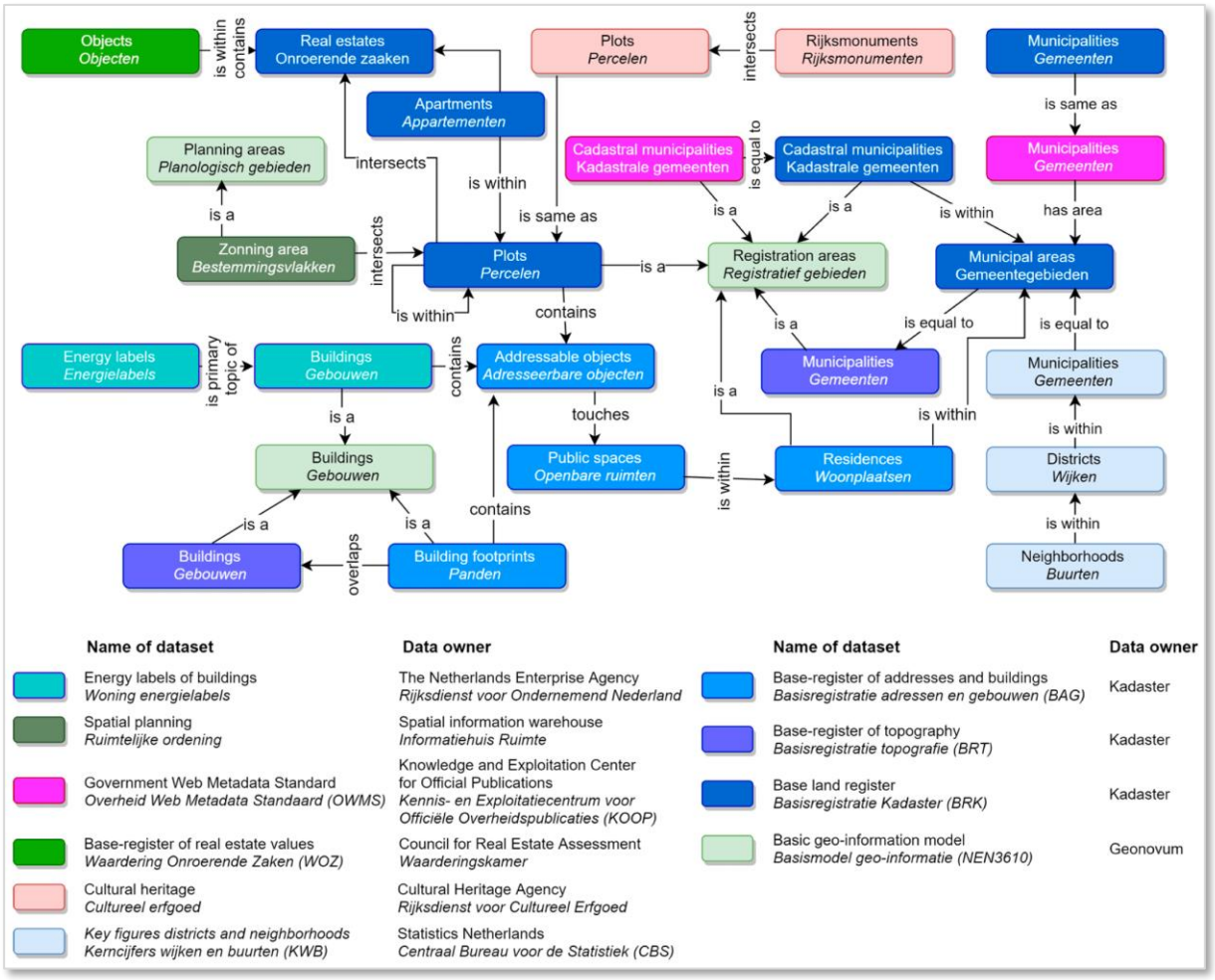

Figure 4: Kadaster Knowledge Graph, © Ronzhin et al., 2019.

The project Linked Cartography of The National Geographic Institute of Spain creates links at the data level between different geographic datasets (topographic, regional, word maps at different scales, highest resolution orthophoto, satellite imagery, etc.) in order to be able to publish them more simply and intuitively for the users. Its general architecture is illustrated in Figure 5.

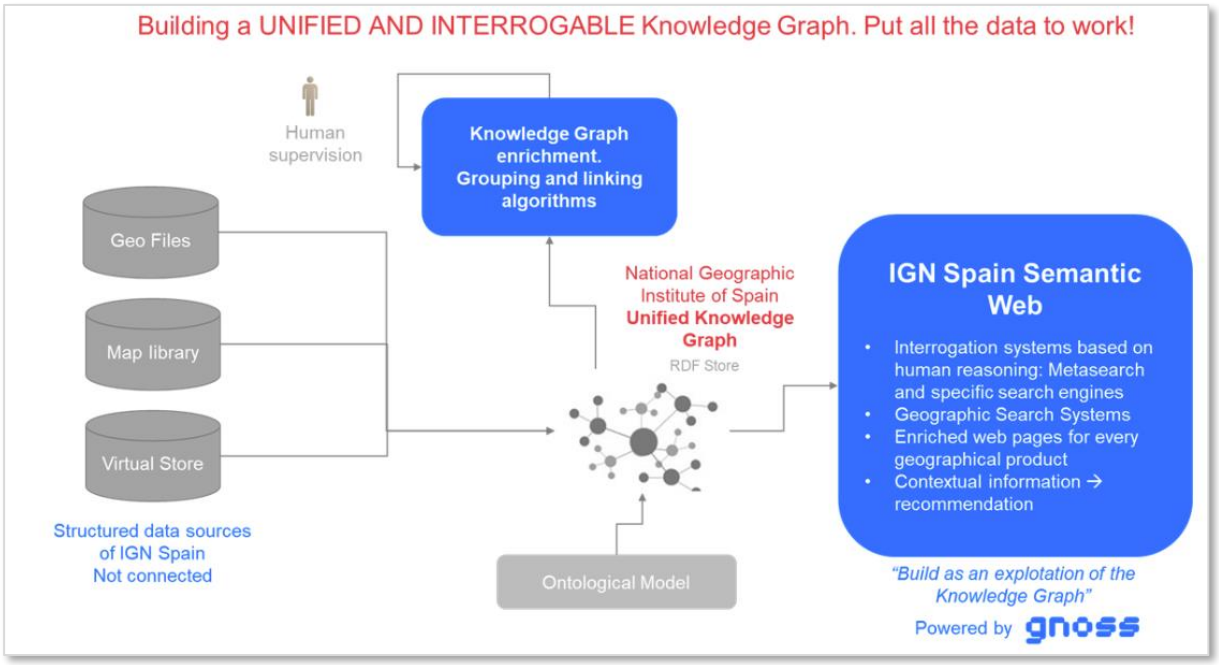

Figure 5: Architecture of the Linked Cartography of Spain project:
a Unified Knowledge Graph is created where geographical data from different scales are interconnected. © GNOSS (after (Bucher et al., 2021))

In Ireland, the linked data technology was also used to interconnect different data sources in a data hub serving different applications dedicated to the management of the COVID-19 crisis, see Figure 6.

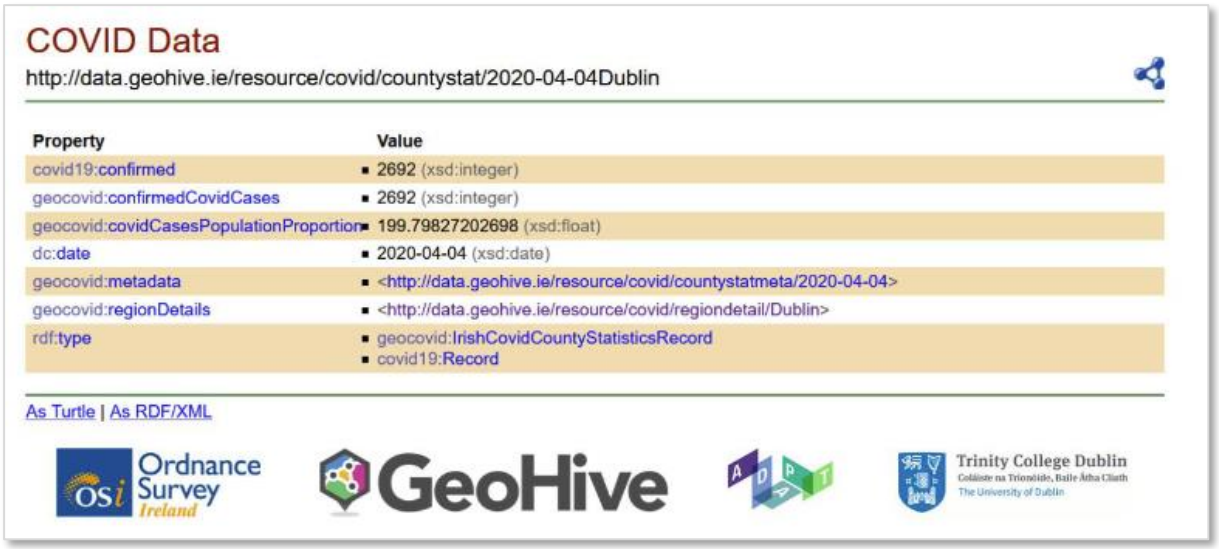

Figure 6: Publishing covid cases through GeoHive as linked data that interconnect different authoritative sources, ©OSI, (after (Bucher et al., 2021)).

The Open European Location Services (OpenELS) project piloted the delivery of the Administrative Units and Named Places datasets as linked data, with Norway, The Netherlands, Finland and Spain providing their data in an RDF format, according to the INSPIRE ontologies created by the Joint Research Centre (JRC) (Rhonzin et al., 2019 b).

The City Orchestra project is funded by a French call for Smart and Sustainable Territories and brings together 18 partners, among which the French Research Center for Buildings (CSTB). It aims to organize and connect open-source tools to ensure the continuity, smoothness, and control of the data processing chain for the ecological transition of territories. The project aims to provide operational solutions for climate change adaptation, air quality improvement, energy efficiency, and responsible land management by efficiently gathering relevant geospatial metadata and data. A Land Cover use case will define a set of specifications (metadata and semantics) to meet the interoperability needs of the consortium.

The ICE ARAGON project, funded by the Government of Aragón, is modeling and integrating different sources into a geographic knowledge graph associated to services about communication network availability planned to be published in 2024. The stakeholders providing real data are distributed across the territory. These currently are either public health or education facilities. The main idea is to integrate data on the internet connection in land cells[7] with other infrastructures (such as the transport network, population, buildings, health and education facilities, etc.) and other data (statistical, socioeconomic) available in existing KGs. For the categorisation of cells, the KG also takes into account the existence of scheduled plans for improving the current network, zones without buildings, differences between theoretical network quality (obtained by models or provided by telecom operators) and real measurements obtained in situ and so on. The integration of data on mobile network availability in a KG allows for a more broad and comprehensive view of the actual situation in a given territory and consequently helps the people in their decision-making processes. Being able to relate this information with other data is a valuable tool for policymakers in their effort to strengthen the least developed zones and to correct problematic imbalances between zones. Having access to up-to-date data on the network capacities is key for the success of this initiative. In order to assure that data on real measurements can be obtained regularly, a solution to task stakeholders is needed, i.e. to request their collaboration to provide a specific measurement. This solution could be grounded on the Knowledge Graph.

2.3 Semantic Web paradigms applied to metadata management

Metadata themselves are a specific category of data. Yet, their production is a tedious task, and the resulting metadata is not always very tractable. Semantic Web (SW) paradigms can be adopted to interconnect metadata during its production as investigated at Dutch Kadaster. A central solution is needed to integrate the different categories of metadata produced and used internally at the Dutch Kadaster organization: structural metadata (that describes a data product, such as definitions of geographical objects or data schemas), administrative metadata (providing data to the associated dashboards), and dataset descriptions. Metadata should be handled as data and an architecture, incl. metadata stewards, is needed to produce and manage metadata as it is the case for data, in general.

Semantic Web (SW) paradigm can also support the engagement of users in the production of metadata. Users can generate relevant metadata related to data even if they are not the data providers. An OGC standard for users generated metadata has been proposed, called Geospatial User Feedback (GUF), illustrated in Figure 7 (Maso et al., 2016). The authors argue that: “Geospatial User Feedback is metadata that is predominantly produced by the consumers of geospatial data products as they use and gain experience with those products. The standard allows for documenting feedback items such as ratings, comments, quality reports, citations, significant events, etc. about the usage of the data. Feedback items can be aggregated in collections and summaries of the collections can also be

[7] Cells are divisions of the territory used for the communication network

described. This standard complements existing metadata conventions whereby documents recording dataset characteristics and production workflows are generated by the creator, publisher, or curator of a data product." Zabala et al. (2021) have extended the GUF model, adapting it to the specific case of research projects, in which users describe their usage as a reproducible workflow, as illustrated on Figure 8. They insist on the need for a solution centralising the expertise acquired by users, and making the GUF statements of users visible to users outside their network.

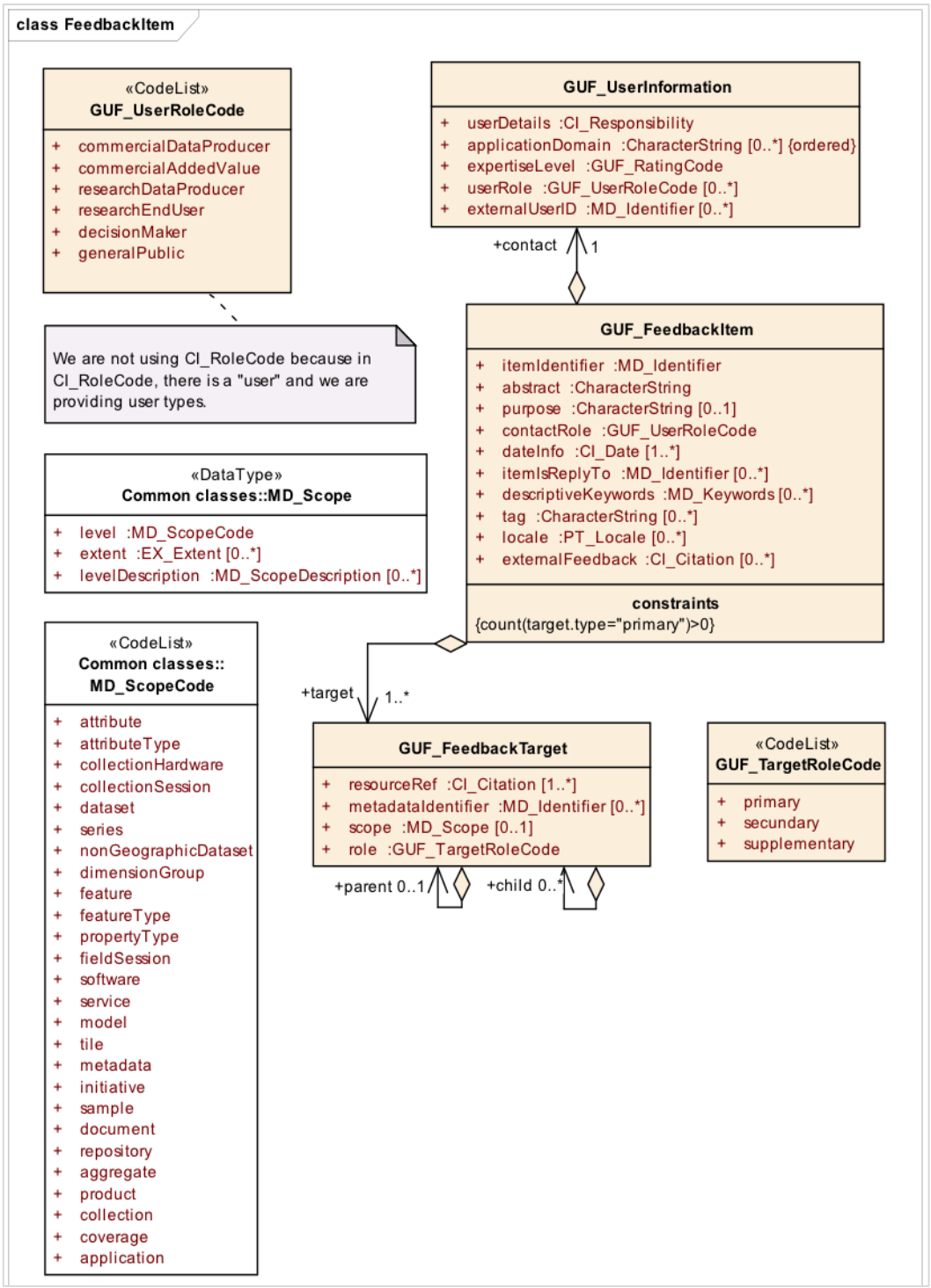

Figure 7: The GUF standard proposed by OGC (Maso et al., 2016) to structure feedback provided by users on datasets and data products to facilitate the management of these feedbacks as relevant metadata.

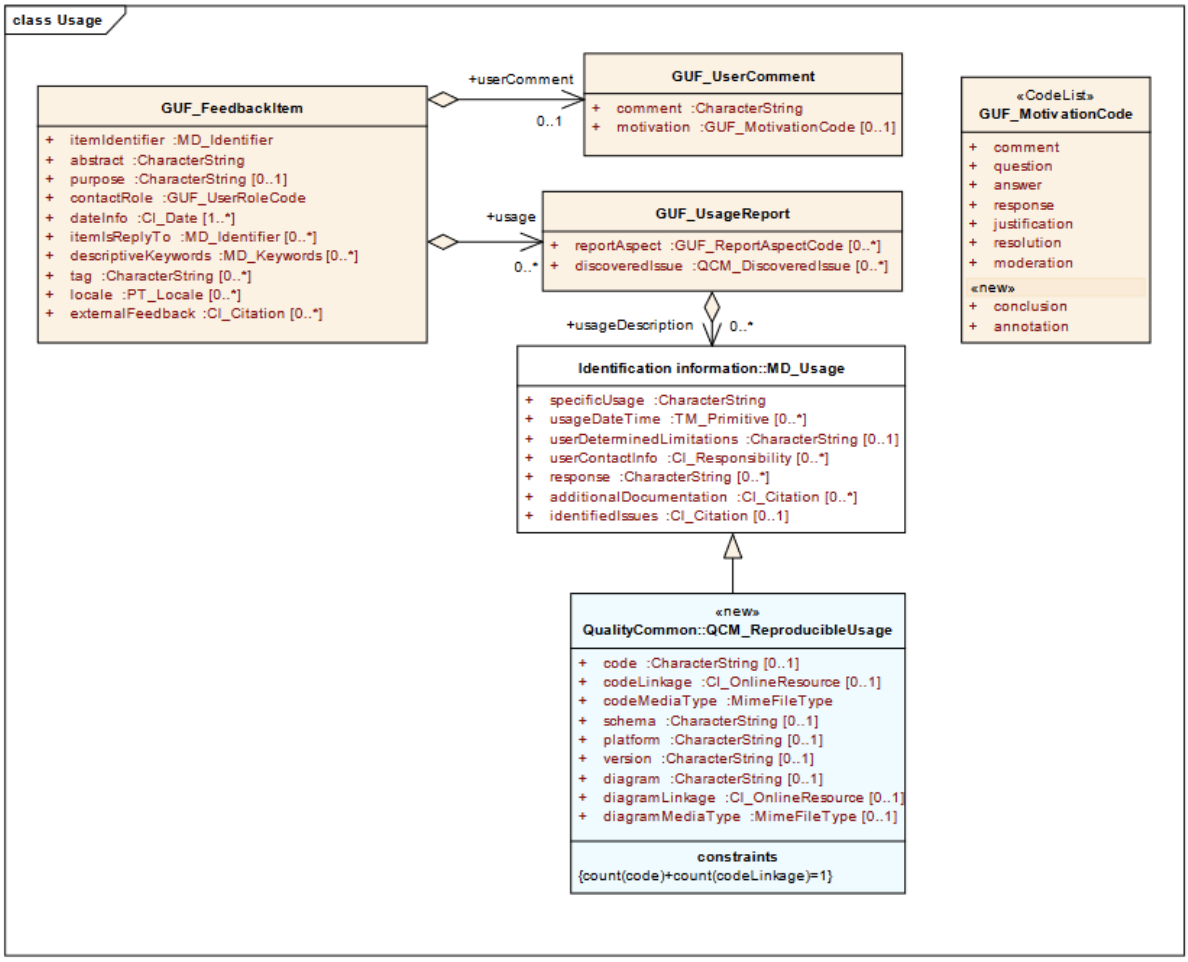

Figure 8: GUF extension proposed by (Zabala et al., 2021) to include more details about the usage and support its reproducibility.

3 EuroSDR KNOWLEDGE GRAPH FOCUS

3.1 Consolidating providers and users expertise about geographical digital assets in Europe

As introduced at the beginning of this report, previous activities of EuroSDR Linked Data group targeted the production of linked metadata. ISO19139 metadata describing existing data products within mapping and cadastral agencies was transformed into RDF metadata. Links between data product nodes were created based on the topic categories and keywords associated to these nodes. This was a draft Knowledge Graph about geo-digital assets in Europe.

However, more details about these data products are available than what is encoded in ISO19139 mandatory metadata elements: for example,detailed provenance information and product specifications. These elements are classically documented by geodata providers. It is now acknowledged that users also gain expertise and knowledge on the geodata products they use and can generate feedback as illustrated on Figure 9.

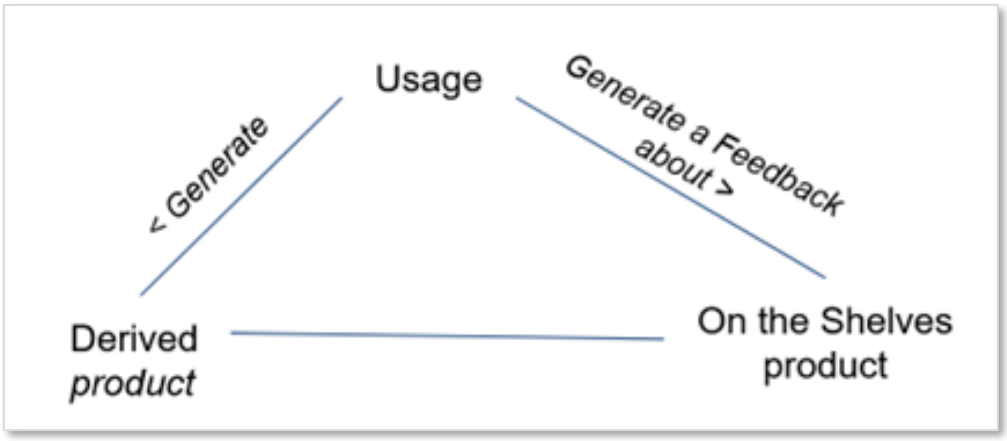

Figure 9: Users of on the shelves (OTS) data products gain expertise about these OTS products that can be shared by generating explicit feedbacks.

There are also richer links between products, than those generated by searching for common keywords. These can be, for example, methods for integrating data from the two different products as illustrated in Figure 10.

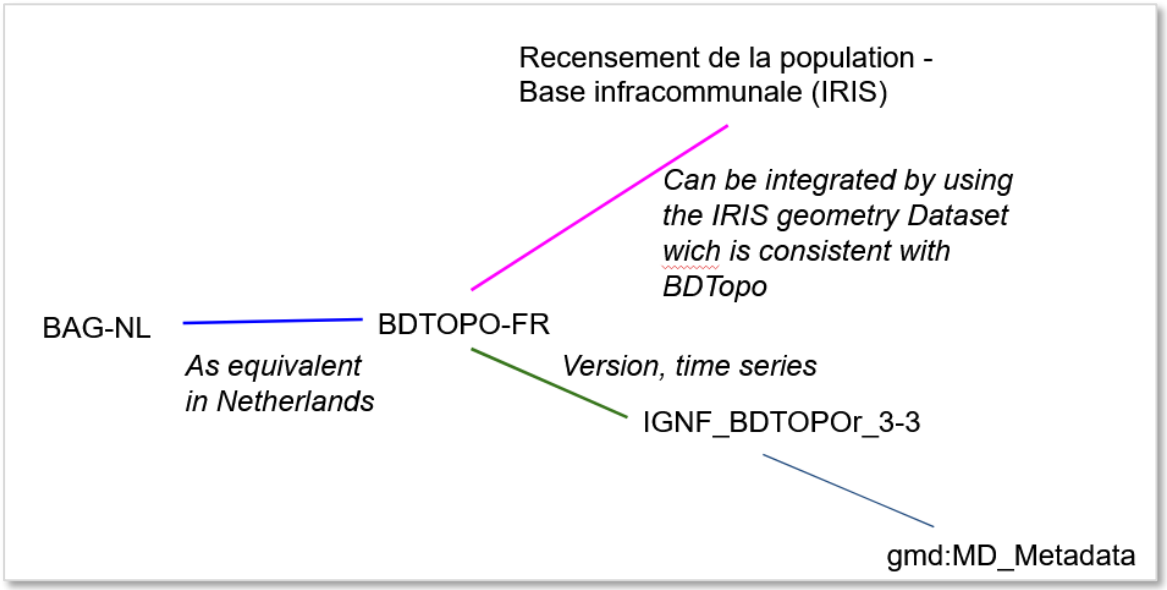

Figure 10: Different kinds of links can be identified between data products:
sister products across countries (see the blue link),
complementary products on a given territory (see pink link),
specific versions of a product (see green link).

In that context, a EuroSDR KG on Geodata products could consolidate expertise and experience related to data products from EuroSDR data provider networks, by supporting the publication, sharing and reuse of Geospatial User Feedbacks (GUF) expressed by users outside mapping agencies as well as possible statements from providers within production lines, incl. expression of links between products.

More specific tasks that could be associated with such a KG are listed below:

- Assist a data user in filling in the citation part of their GUF, in the cases for which the referred data product has already been referenced in the EuroSDR KG.
- Assist a data provider in enriching the KG by publishing additional properties related to their products and links with other products.
- Assist a data user in retrieving relevant feedback items from other users and data providers.
- Assist a data provider in retrieving relevant feedback items from other users and from data providers.

3.2 Prototype's focus

3.2.1 General considerations

To foster the prototype development, two more specific context have been identified where the consolidation of users expertise is needed and where EuroSDR LD group could participate. The first context is to consolidate expertise related to the data documentation and quality and to metadata. The second context is to share expertise related to the integration of different data products together to generate a new ad hoc product.

The development of EuroSDR KG prototype is targeting the following activities:

- Establishing a core KG structure, reusing fragments from past activities.
- Specifying the necessary APIs, which support the production of RDF data for the KG: to register new data products, new software, express geospatial user feedback, enrich description of existing product.
- Specifying the necessary APIs to support querying the KG.

The general architecture, in which the prototype is embedded, is schematised in Figure 11. The project will concentrate on setting up and feeding the github repository as well as specifying protocols for possible interactions with other components, in particularly the working environments of targeted users and targeted contributors. Plugins developments initially envisioned have been postponed due to the difficulty of recruiting a master student with the required engineering skills.

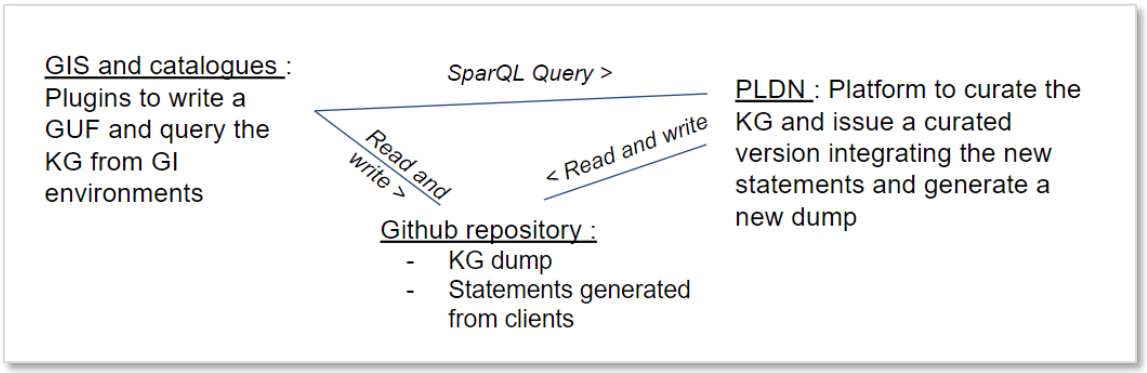

Figure 11: General architecture associated with EuroSDR KG.

We adopt github as a new platform for our group and project, https://github.com/EuroSDR-LDG.

3.2.2 Sharing experiences on a specific product quality and documentation

Consolidating users alerts on national geoportals

This use case consists in evaluating the potential of KG technologies to better structure and organize users' feedback expressed through NMCAs geoportals:

- messages on forum associated with services platforms,
- alerts raised through Web interfaces for users to inform the NMCA that some data needs update.

A set of alerts will be selected and manually analyzed to identify core named entities and terms that would be needed for an automatic extraction of structured statements from these alerts.

Sharing expertise on Reading Detailed Documentation for Diachronic analysis

In this scenario, the user would like to share with others his experience in identifying and gathering fragments of metadata among the rich documentation of IGN topographic products (BDTopo(r)) that correspond to his specific application.

The specific application is to use the French, German and British topographic building data on the past decade to study the evolution of buildings. Scientists need to identify evolutions of the product specifications that could bias the study. In the case of French BDTopop®, these evolutions are documented in a text-based documentation accessible from the French geoportal geoservices.ign.fr. . These documents are detailed descriptions of the products, usually specific to a given version, and sometimes summaries of evolutions between versions. Within a detailed product description there can also be explanations of the evolutions of the product specification with respect to a category of object, for example that BDTopo® selection threshold for buildings was initially 50m2 and that after unifying the product with the french cadastral product, there is no more threshold. The ISO19139 metadata also contains relevant statements like the mention that all features have a stable identifier, yet not specifying since when they do (highlighted fragment in yellow). Last, specific metadata are published to describe the production process, like the date of the aerial photograph campaign that was used to update the data of a given area. All these fragments need to be extracted from the documentation and gathered to guide the diachronic analysis.

The experience will consist in supporting the citation of the given data product (BDTopo®, OS MasterMap®, ATKIS®) as well as the citation of each documentation sources (text document, ISO19139 metadata, csv or shape files), and quoting the fragments that are relevant to the application at stake, diachronic analysis.

The same use case will also support the provision of additional expertise from production teams, reacting to the initial feedback by suggesting another relevant documentation source or directly writing down some information that was left implicit.

3.2.3 Feedback and experiences on the integration of data from different products

Integrating Statistical Data with Topographic Data

In this scenario, a participant would like to describe the development of a prototype service utilising the OGC API - Joins draft standard (OGC, 2023) for data integration. The service has been implemented in the GeoE3 project and it integrates statistical data from the Statistics Finland's PxWeb API with Finnish municipalities and regional datasets (Latvala et. al., 2022). The data integration is performed by creating a new dataset from input datasets. The integrated dataset contains all the available geospatial features, together with the selected statistical data items that have been successfully joined with the geospatial dataset using the selected identifier fields.

The experience will consist in writing this experience in an informative way so that another user can evaluate the feasibility of reproducing it. EuroSDR KG should support the citation of data sets and APIs through unambiguous identifiers. The users also need to express that the integration process requires that the identifier fields in the geospatial and statistical dataset contain the same values, and to report on any problems encountered in the data integration process.

Integrating data to produce Digital Twins or KGs of geographical entities

Some participants to EuroSDR Linked Data project also are involved in the City Orchestra project or in the ICE ARAGON project. They will be sharing feedback on datasources and datasets manipulated during their projects. One category of feedback we wish to focus on is the use of multiple data sources and their integration into a new dataset. This can be, for example, integrating different data related to buildings, parcels, addresses, population data from statistical survey, and possibly a Building Information Model (BIM) in the case of City Orchestra.

In these scenari, most of the GUF will be written as free text. The assistance from the EuroSDR KG would be supporting the citation of the data sources, giving information on how the sources were integrated, as well as registering the new derived data product and its properties, the schema used and the fact it is a linked dataset.

4 CONCLUSION

The purpose of this document was to present the focus of EuroSDR Knowledge Graph prototype from EuroSDR Linked Data project.

Sections 1 and 2 describe the context of this prototype: the stake of sharing and reusing geographical data and the relevance of Semantic Web in that matter.

Improving data sharing and reuse is a hot topic to benefit from the growing amount of data sources and advances in integration technologies to inform sustainable development of our territories. For NMCAs, this general issue of data sharing and reuse correspond to different more specific stakes:

- Discoverability of their products from potential users,
- Monitoring of the effective usage of the data,

- Discoverability of data for their own activity: either internally or from external sources,
- Metadata management.

Semantic Web (SW) technologies, in particular Linked Data and Knowledge Graphs (KG), present opportunities to meet these stakes and are already adopted in some projects briefly presented in this report. Some approaches concentrate on designing linked metadata and interconnect them with domain KG to produce cross domain catalogs, or to develop Query Answering capacities. Other approaches apply SW technologies to the data level, either to create linksets or to transform data into linked data.

Section 3 presents the more specific focus of EuroSDR Knowledge Graph : to consolidate and share expertise and experience related to geodata products in Europe. These include products from Publicly Mandated Authorities (PMA) or other stakeholders (e.g. DBpedia, OSM) and also ad hoc products derived for specific applications. It will support the publication, sharing and reuse of Geospatial User Feedbacks (GUF) expressed by “users” outside mapping agencies or “data experts” within production lines.

The purpose of EuroSDR KG prototype is to demonstrate the expected benefit from the KG on geodata products as well as the costs to achieve it. It should serve as the stepping stone towards a larger and more ambitious project.

Consequently, EuroSDR KG prototype has been specified in this report based on real needs:

- Consolidating users alerts on NMCAs geoportals
- Sharing experience related to gathering metadata fragments relevant to an application at stake, here, diachronic analysis of building based on topographic building data.
- Sharing experience related to the integration of different sources to generate a knowledge graph at the level of data.
- Sharing experience related to integrating topographic and statistical data.

This is a KG at the level of metadata, which can be useful to inform the production of KG at the level of data in further work.

To adapt to the architecture requirements, the group is shifting its platform from the former wiki, geometadatalabs.eu, to github: https://github.com/EuroSDR-LDG. Such a platform is more adapted to the requirements of our group, to share vocabularies and fragments of data and codes. An important aspect of our project is its openness allowing anyone to join at any time. We therefore warmly invite readers to join the project.

Acknowledgements

This work was supported by EuroSDR in the context of the Linked Data project. The authors wish to warmly thank Conor Cahalane from Maynooth University and Joep Crompvoet from KU Leuven for their thorough and constructive reviews.

References

Archer, P., Blower, J., Calegari, N., Cochrane, B., Cox, S., Daoust, F., Harth, A., Hurst, I., van Leeuwen, B., Lieberman, J., Little, C., Mabbett, A., Parslow, P., Parsons, E., Perego, A., Portele, C., Roberts, B., G. Svensson, L., Taylor, K., Wilde, E., Atkinson, R., Lewis, E., 2023, *Spatial Data on the Web Best Practices*, OGC/W3C Group Draft Note, Tandy, Vans den Brink, Barnaghi, Homburg (eds), https://www.w3.org/TR/2023/DNOTE-sdw-bp-20230919/ OGC Document Number OGC 15-107

Bucher, B., Tiainen, E., Ellett von Brasch, T., Janssen, P., Kotzinos, D., Čeh, M., Rijsdijk, M. Folmer, E., van Damme, M.-D., Zhral, M., 2020, *Conciliating Perspectives from Mapping Agencies and Web of Data on Successful European SDIs: Toward a European Geographic Knowledge Graph*, in ISPRS International Journal of Geo-Information, 9 (2), pp.62, https://www.mdpi.com/2220-9964/9/2/62

Bucher, B., Folmer, E., Brennan, R., Beek, W., Hbeich, E., Würriehausen, F., Rowland, L., Alonso Maturana, R., Alvarado, E., Buyle, R., Di Donato, P., 2021, *Spatial Linked Data in Europe: Report from Spatial Linked Data Sessions at Knowledge Graph in Action*, EuroSDR Official Publication No.73, https://eurosdr.net/publications/official-publication-no-73-2021

Bucher, B., Grudzien, M., Delattre, N., Escriu Paradell, J., Folmer, E., Garrone, A., Kügeler, A., Lopez, A., Parsons, E., Perego, A., Pilar, J., Reini, J., Reuter, H.-I., Saligoe-Simmel, J., Schneider, M., Ticheler, J., 2023, *Geodata Discoverability*, EuroSDR Official Workshop Report, 35 pages, https://eurosdr.net/publications/workshop-report-geodata-discoverability-2022

Heath, T., Bizer, C., 2011, *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., Kirrane, S., Emilio Labra Gayo, J., Navigli, R., Neumaier, S., Ngonga Ngomo, A.-C., Polleres, A., M. Rashid, S., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., Zimmermann, A., 2021, *Knowledge Graphs*. ACM Comput. Surv. 54, 4, Article 71, 37 pages. https://doi.org/10.1145/3447772

Latvala, P., Huuhko, K., Kokkonen, M., 2022, *Integrating Statistical Databases with Geospatial Datasets*, Abstr. Int. Cartogr. Assoc., 5, 121, https://doi.org/10.5194/ica-abs-5-121-2022

Masó, J.; Bastin, L.; Thum, S., Nüst, D., 2023, *OGC Geospatial User Feedback Standard: Conceptual Model*, Ver. 1.0, 15-097r1; Open Geospatial Consortium: Wayland, MA, USA, 2016, OGC. OGC API - Joins. https://github.com/opengeospatial/ogcapi-joins

Masser, I., Crompvoet, J., 2018, *Qualitative monitoring of information infrastructures: A case study of INSPIRE*, in Environment and Planning B: Urban Analytics and City Science, Vol 45, Issue 2, https://doi.org/10.1177/0265813516675871

Masson, V., Bocher, E., Bucher, B., Chitu, Z., Christophe, S., Fortelius, C., Hamdi, R., Lemonsu, A., Perrels, A., Van Schaeybroeck, B., Wichers Schreur, B., Velea, L., Beddar, Y., Calvet, J.-C., Delcloo, A., Druel, A., Duchêne, F., Dumas, G., Dumitrescu, A., Gautier, J., Goret, M., Horttanainen, M., Kouznetsov, R., Le Bris, A., Lecorre, S., Le Roy, B., Palamarchuk, Y., Petit, G., Ruuhela, R., Saranko, O., Sofiev, M., Siljamo, P., Van de Vyver, H., van Velthoven, P., Votsis, A., 2020, *The Urban Climate Services URCLIM project, CLIMATE SERVICES*. vol. 20, ISSN 2405-8807, https://doi.org/10.1016/j.cliser.2020.100194

Ronzhin, S., Folmer, E., Maria, P., Brattinga, M., Beek, W., Lemmens, R., van't Veer, R., 2019, *Kadaster Knowledge Graph: Beyond the Fifth Star of Open Data*. Information 2019, 10, 310. https://doi.org/10.3390/info10100310

Ronzhin, S., Folmer, E.J., Lemmens, R.L., Mellum, R., Brasch, T.E., Martin, E.P., Romero, E.L., Kytö, S., Hietanen, E., & Latvala, P., 2019b, *Next Generation of Spatial Data Infrastructure: Lessons from Linked Data implementations across Europe*. Int. J. Spatial Data Infrastructures Res., 14.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al., 2016, *The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data 3, 160018 https://doi.org/10.1038/sdata.2016.18

Zabala, A., Masó, J., Bastin, L., Giuliani, G.; Pons, X., 2021, *Geospatial User Feedback: How to Raise Users' Voices and Collectively Build Knowledge at the Same Time*. ISPRS Int. J. Geo-Inf. 10, 141. https://doi.org/10.3390/ijgi10030141

Zrhal, M., Bucher, B., Van Damme, M.-D., Hamdi, F., 2021, *Spatial dataset search: Building a dedicated knowledge graph*, in proceedings of the 24th AGILE Conference on Geographic Information Science, AGILE GI Science Series, https://doi.org/10.5194/agile-giss-2-43-2021