EuroSDR

European Spatial Data Research

February 2021

Spatial Linked Data in Europe: Report from Spatial Linked Data Session at Knowledge Graph in Action

October 6th 2020 - On-Line Conference

Bénédicte Bucher, Erwin Folmer, Rob Brennan,
Wouter Beek, Elio Hbeich, Falk Würriehausen,
Lexi Rowland, Ricardo Alonso Maturana,
Elena Alvarado, Raf Buyle, Pasquale Di Donato

Official Publication Nº 73

Published by EuroSDR

COMMISSION CHAIRPERSONS:

Data Acquisition: Jon Mills, United Kingdom
Modelling and Processing: Norbert Haala, Germany
Updating and Integration: Martijn Rijsdijk, The Netherlands
Information Usage: Bénédicte Bucher, France
Business Models and Operation: Joep Crompvoets, Belgium
Knowledge Transfer: Markéta Potůčková, Czech Republic

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, Erwin Folmer, Rob Brennan, Wouter Beek, Elio Hbeich, Falk Würriehausen, Lexi Rowland, Ricardo Alonso Maturana, Elena Alvarado, Raf Buyle, Pasquale Di Donato

“Spatial Linked Data in Europe:
Report from Spatial Linked Data Sessions at Knowledge Graph in Action”

October 6th 2020 - On-Line Conference

Table of contents

SPATIAL LINKED DATA IN EUROPE: REPORT FROM SPATIAL LINKED DATA SESSIONS AT KNOWLEDGE GRAPH IN ACTION

With 11 figures and 1 table

Bénédicte Bucher [a,b], Erwin Folmer [c,d], Rob Brennan [e], Wouter Beek [f],
Elio Hbeich [g], Falk Würriehausen [h], Lexi Rowland [c],
Ricardo Alonso Maturana [i], Elena Alvarado [i], Raf Buyle [j,k], Pasquale Di Donato [l]

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

[b] University Gustave Eiffel
France

[c] Kadaster
Enschede, The Netherlands

[d] University of Twente
Enschede, The Netherlands

[e] ADAPT, Dublin City University
Dublin, Ireland

[f] Triply B.V.
Amsterdam, The Netherlands

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

[h] Federeal Agency for Cartography and Geodesy –BKG
Frankfurt am Main, Germany

[i] GNOSS
La Rioja, Spain

[j] Flanders Information Agency
Ghent, Belgium

[k] Ghent University
Ghent, Belgium

[l] Federal Office of Topography swisstopo
Wabern, Switzerland

Index of Figures

Index of Tables

1 INTRODUCTION

In 2020, the Knowledge Graph in Action (KGiA) online conference was organized as a joint event gathering three annual events with a common interest on producing, consolidating data, and supporting their joint reuse and different specific focuses within this common interest: the DBpedia day which more specifically focuses on advancing DBpedia, the EuroSDR Spatial Linked Data day which more specifically focuses on spatial linked data, and the EuroSDR VGI event which more specifically focuses on volunteered geographic information.

The event was organized around distinct parallel sessions dedicated to each event and joint plenary sessions. During plenary sessions, keynotes related to the modelling and the usage of spatial knowledge, in particular in the form of knowledge graphs, at the junction of these communities. Carsten Hoyer Click from the German Aerospace Center presented the design of a development of a distributed data infrastructure for energy systems analysis. Semantic Web techniques are used to interconnect data from different sources and prepare the integrated data layers needed for energy models. Peter Mooney from Maynooth University presented opportunities for more collaboration and geo-information integration between volunteered geographical information, the governmental agencies and the geospatial research communities. He insists on the complexity of data integration, which is always present even when flowcharts hide this complexity and on the semantics aspect being the more difficult to solve. Here the exploration of machine learning and artificial intelligence are the dominant trend. Marinos Kavouras from the National Technical University of Athens extended upon the need for our society to develop competences to interpret all the data available, in big quantities, to make sense of complex phenomena. He argues that space has been one of the strongest pivotal notions in semantically linking all kinds of data. Developing geospatial literacy skills is needed to empower people with a modern cartographic language, an indispensable communication and cognitive tool. Krzystof Janowicz from the University of California presented the application of knowledge graphs to address challenges at the interface between humans and their environment like for example crisis management. The information currently provided to end users is based on the integration of highly heterogeneous data from different fields of expertise and can lead to misinterpretation. Knowledge graphs and their technologies offer perspectives and lots of challenges still ahead to make data AI-ready at the level of individual statements instead of merely offering access to datasets, to provide additional contextual background information.

The rest of this report concerns presentations and exchanges that took place during the EuroSDR Spatial Linked Data sessions. EuroSDR is a not for profit association established since 1953 for the purpose of applied research and innovation in spatial data provision, distribution and usage in Europe. It gathers national mapping agencies, research institutes, universities and industries. Its activity on Linked data has two main objectives : 1) assessing the value of this technology in addressing current challenges in spatial data provision, distribution and exploitation, 2) identifying new needs for spatial data provision and distribution that have emerged with this technology. This activity started in 2015 and is grounded on big events -like the KGiA conference-, smaller working sessions, and since 2019 a technical group. EuroSDR LD group gathers participants with an interest in Spatial Linked Data (SLD). SLD can be characterized as a domain of applied research and innovation at the overlap between Linked Data and spatial data. Its finality is data production, sharing and reuse on the Web to support decisions with a geographical characteristic. Space is an important dimension to interconnect different information and achieve the Linked Data vision, for example to valorise linked data of different domains if any spatial footprints can be added to associate them with a geographical context or to detect possible connections between different data not connected otherwise. Vice versa, graph based models are promising approaches to address some unsolved issues in spatial data infrastructures.

The section “National presentations” reports on updates presented by different agencies or partners on latest developments, focusing on a given territory. These developments are either in a prototype stage or were presented as fully operational applications.

The remaining sections report on more technology oriented presentations.

The section entitled "Interfacing more users with data and related technologies" present results and approaches oriented on the appropriation of data by potential users, despite possible silos created by the complexity of data technologies, including linked data, was addressed in several presentations. The self-service GIS vision presented by the Kadaster is to support the querying and exploitation of complex data by more users beyond the limited Geomatics Community. The tools developed by Triply, in particular a wizard, focus on giving access to the potential of Linked data to users who are no LD specialists thanks to user oriented interfaces. Besides, a well known usage of Knowledge Graphs is to improve user access to resources -as on Amazon, AirBnB, Google and other platforms, based on the modelling in a knowledge graph of important knowledge related to the resources and also related to the usage. This can be applied in particular to specific resources: the data themselves. The discovery of "fitted for use" datasets, especially spatial datasets is a pending issue given the wide range of users on the one hand, and the difficulty to broker and compare datasets potentially relevant on the other hand. A new EuroSDR initiative targeting the design of an open European Knowledge Graph of geographical digital assets was presented. It consists of the collaborative creation of an open Knowledge Graph about digital assets in Europe, based on the EuroSDR LD Group sandbox and EuroSDR community as a whole.

The last section reported on GeoSPARQL focused presentations. A key technology associated with Linked Data and the Knowledge Graph is GeoSPARQL. One presentation focused on requirements from the domain of buildings and on the type of spatial queries that should be addressed to 3D linked data. Another presentation concerned the GeoSPARQL benchmark on the EuroSDR sandbox.

2 NATIONAL PRESENTATIONS ON SPATIAL LINKED DATA ACTIVITIES

2.1 Kadaster update (Netherlands)

The Netherlands has a long history of publishing key government registers as Linked Open Data, many of which, including Topography, Addresses and Building, and Cadastral Parcels, have existed for many years. In publishing these, Kadaster has taken a traditional approach in transforming a siloed dataset into a siloed linked data set. This work amounts to more or less a change in syntax and while this does include some benefits such as transparent semantics, the most obvious value of doing such a transformation was not achieved in approaching transformation in this way, leaving us with room to improve. This notion, in combination with the fact that a lot of progress has been made in the way of software support for publishing linked data, has led to new activities in three areas within Kadaster as outlined below.

1. Tools showcasing the value of Linked Data

   From a user perspective, Linked Data is a relatively new concept and there is a recognition that without tools to showcase the potential that this has, as well as make interaction easier, the learning curve remains steep and accessibility is an issue. As such, Kadaster has invested and continues to invest in the availability of tools including Data Stories, facet browser, SPARQL tools and self-service applications which support a wider range of users in their accessibility and interaction with our datasets. The aim is to provide for the needs of users ranging from members of the public with little technical experience to developers and programmers.

2. Additional Focus on Knowledge Graph

   Providing siloed Linked Data is just a starting point and does not provide that many user benefits. Linked Data suffers from network effects where the user value of the technology will only increase with a growing number of linked datasets. The new focus, then, is also on connecting the silos and offering a Knowledge Graph made up with this connected data that can be browsed and queried in a user-friendly way. Our starting point is connecting Kadaster datasets together to form the Kadaster Knowledge Graph. It is possible that this Graph can be used as the starting point for larger Knowledge Graph initiatives where Kadaster data is linked to other open government and community datasets. An example of this is the Dutch Knowledge Graph initiative by the DBPedia association.

3. New Linked Data Stack

   Kadaster's publication stack relating to Linked Data, which primarily contained components from relational database paradigm, has been replaced by a more native Linked Data stack. This new stack encompasses and manages the whole lifecycle of linked data, from storing, integration and linking to publishing, analysing and visualisation; improving the performance gaps that were often seen in the earlier publication stack. This makes the previous publications more compatible and, as such, the initial notion that a combined stack for multiple publication formats would work in practice has undergone some change wherein the belief now is that a specific native Linked Data stack is a better fit. In line with this new belief, the publication of Linked Data can be done at low-cost with minimal Extract-Transform-Load (ETL) effort but remains fully (linked data) standards-compliant. This new stack is now used for the publication of the fourth key register as Linked Data, the BGT, but also highlights that we need to update our legacy with regards to Linked Data, that is to say the first generation of spatial linked data published by Kadaster.

4. The development of "Self-Service GIS"

   The availability of tools, the knowledge graph, as well as the new linked data stack, makes it possible to dream about our next step. This next step in our vision is to achieve Self-Service GIS, as a variant of the typical functionality found within traditional GIS interfaces, is to provide non-expert users with the opportunity to perform simple geospatial analysis and querying tasks on the web and, where possible, real-time data. This is presented further in this report in a dedicated section.

5. Building a national KG for the Netherlands

   Kadaster has contributed as a source of authoritative data to the development of the Dutch National Knowledge Graph (DNKG), an initiative undertaken by DBPedia. The benefits for data providers of developing such a Graph include improved data quality, increased traffic and enrichment through integration of data sources.

2.2 Kartverket update (Norway)

Kartverket began its Linked Open Data (LOD) work in 2017 after a request from the digitalisation directorate for a solution to indirect geospatial referencing of the records in the national data catalogue. This led to the national administrative units dataset being made available as LOD through an openly accessible rest API (https://rdf.kartverket.no/api/1.0#/adminstrative_unit). This API has now been integrated into the metadata creation and registration process within both the Norwegian national data catalogue as well as the Norwegian Mapping Authorities own geospatial portal catalogue, allowing users to tag metadata records with the persistent URI's (a skos:preflabel is shown to the user in html view) of the applicable administrative unit. There are plans in 2021 to consider adding the national place names dataset to Kartverkets LOD offering, and this might be especially important to other organisations wishing to geocode indirectly using a higher resolution dataset than administrative units.

Within Norway as a whole, the ongoing work is somewhat fragmented, with the main coordinated work focused on delivery of metadata according to the DCAT standard. But as well as the work within the National Mapping Authority, research work is ongoing at SINTEF and the Norwegian Road Authority are currently looking at delivering national road data as LOD.

2.3 BKG update (Germany)

In 2020, the Federal Agency for Cartography and Geodesy (BKG) initiated activities on linked geodata as part of a spatial data infrastructure project in Germany. The focus of the project is on the implementation of an intelligent platform that supports the integration of various linked open data sources (e.g. Wikidata, LinkedGeodata.org) but also existing standards-based data (e.g. for INSPIRE). It is a matter of integrating all available data attributes, format specifications relevant for import or export and the development of new data services that are only possible through semantic data integration.

In order to provide a prototype of linked data in the BKG, various components had to be developed that were to enable the provision of linked data (SPARQL Endpoint), the conversion of various geoformats to Resource Description Framework (RDF) and the exploration and reconversion of linked data into geoformats for their provision. All components together form the basis for a linked data infrastructure in the BKG and can be integrated into various work processes.

A basic requirement for linked open data applications is to describe each resource of each triplet with a URI. It is recommended that these URIs be made resolvable in the web browser so that users can understand the data behind the URI and its relationships to other related data resources. This function is fulfilled in the project by the tool GeoPubby (https://github.com/i3mainz/geopubby), an extension of an earlier version of the DBPedia tool Pubby (https://github.com/cygri/pubby). In contrast to Pubby, GeoPubby (see figure 1) implements a fuzzy search, a map application and offers the possibility to display the current resource in a spatial data format such as Download GeoJSON or GML. In the future, other vector formats such as GeoJSON-LD and the various EPSG coordinate reference systems will also be supported.

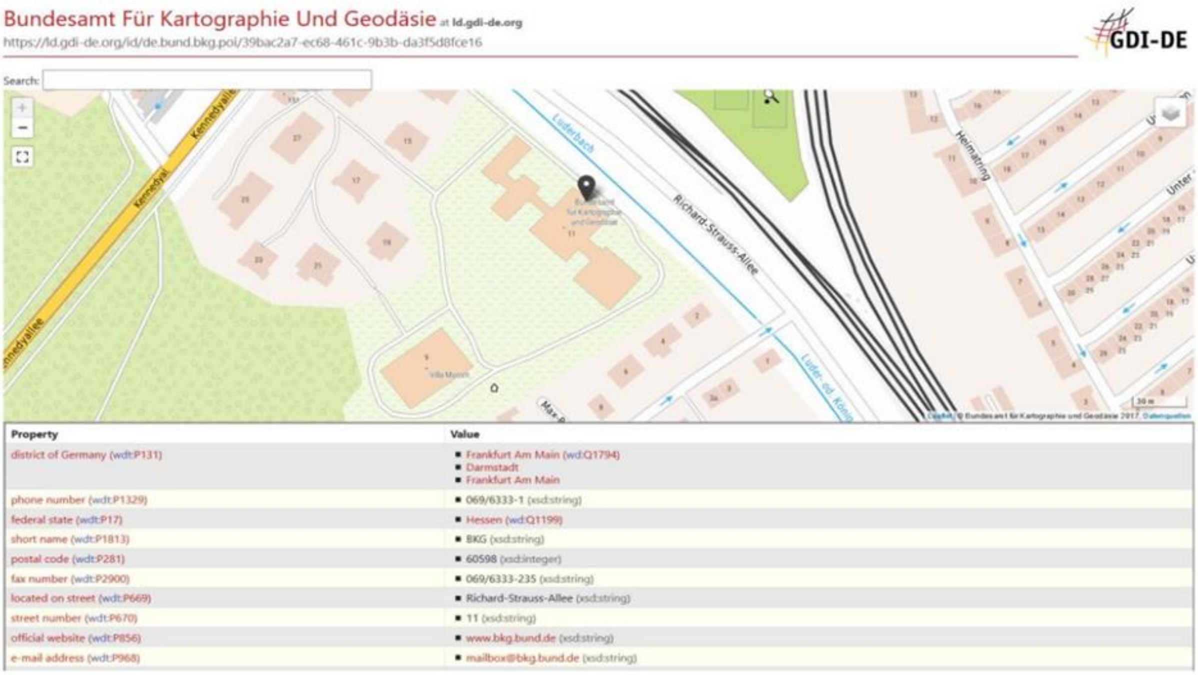


Figure 1: GeoPubby interface

In addition, as a further step, the annotation of parameters for the data quality should also be considered when the data is integrated into the system. A corresponding extension of the application is planned in the BKG for the year 2021.

2.4 OSi update (Ireland)

Ordnance Survey Ireland (OSi), the Irish national mapping agency, have been publishing spatial Linked Open Data since 2016. This has been grounded on a collaboration with Dublin City University and Trinity College Dublin acting together within the Science Foundation Ireland ADAPT Centre for Digital Media Technology. The first trial data published was a political boundary dataset with three available polygon resolutions of 20m, 50m and 100m. A custom version of Pubby was developed as a Geo Linked Data frontend that uses the OSi national map image to visualise the polygons and to browse the data. Data dumps are also available. See data.geohive.ie as an entry point to all the data.

Initial versions of the data.geohive.ie data portal were served by prototype platforms. Internally OSi has since moved onto production systems based on the Oracle Spatial and Graph platform. This has allowed for the development of mappings from relational databases to RDF datasets, using a W3C standard for the expression of such mappings called the R2RML[1] language. These mappings dynamically create RDF "views" of the OSi's National Map (Prime2) data. Prime2 is much greater in extent than the open data published as it tracks over 50 million spatial objects in Ireland with extensive attributes, metadata, versioning and provenance. In 2019 OSi made available a snapshot of Linked Open Data describing all the buildings in County Galway, Ireland (our third city). This was derived from Prime2 using R2RML mappings and includes building identifiers. These buildings IRIs are the basis of a national Unique Geographic Identifiers (UGIs) scheme for Ireland as Linked Data as all 50 million spatial objects in Prime2 can be encoded in this scheme. In addition it has fed into collaborations on linking Buildings Information Models into the National Map. International collaboration in this space is ongoing in aW3C community group[2].

In 2019-20 the collaboration has shifted from the external publication of Linked Data to enabling data governance of the whole OSi data production pipeline through RDF-based data governance metadata. First we focused on data quality metrics for geo-linked data and then integrating quality measurements made at different points in the production pipeline. This required development of a set of mappings between different quality standards such as ISO 19157 and ISO 25012 as different tools generated data according to the different models of quality dimensions in each standard. Our baseline tool for Linked Data is Luzzu[3] and for geospatial data validation is 1Spatial 1Integrate (see figure 2). Through the use of R2RML mapping rules and a set of custom metric definitions it was possible to uplift the 1Spatial 1Integrate outputs into a Dataset Quality Vocabulary (daQ)[4] representation of the metric observations. Then it was possible to build an end to end data quality dashboard showing quality observations at various stages in the OSi data production pipeline and to visualise the results according to the user's preferred data quality standards (ISO 19157 or ISO 25012). This infrastructure supported the publication of data useful to manage the COVID crisis (see figure 3). The next steps will be to incorporate the full range of data governance metadata (catalogs, provenance) with the data quality measurements to enable root cause analysis and wider data governance controls.

[1] https://www.w3.org/TR/r2rml/
[2] https://www.w3.org/community/lbd/
[3] https://eis-bonn.github.io/Luzzu/
[4] http://theme-e.adaptcentre.ie/daq/daq.html

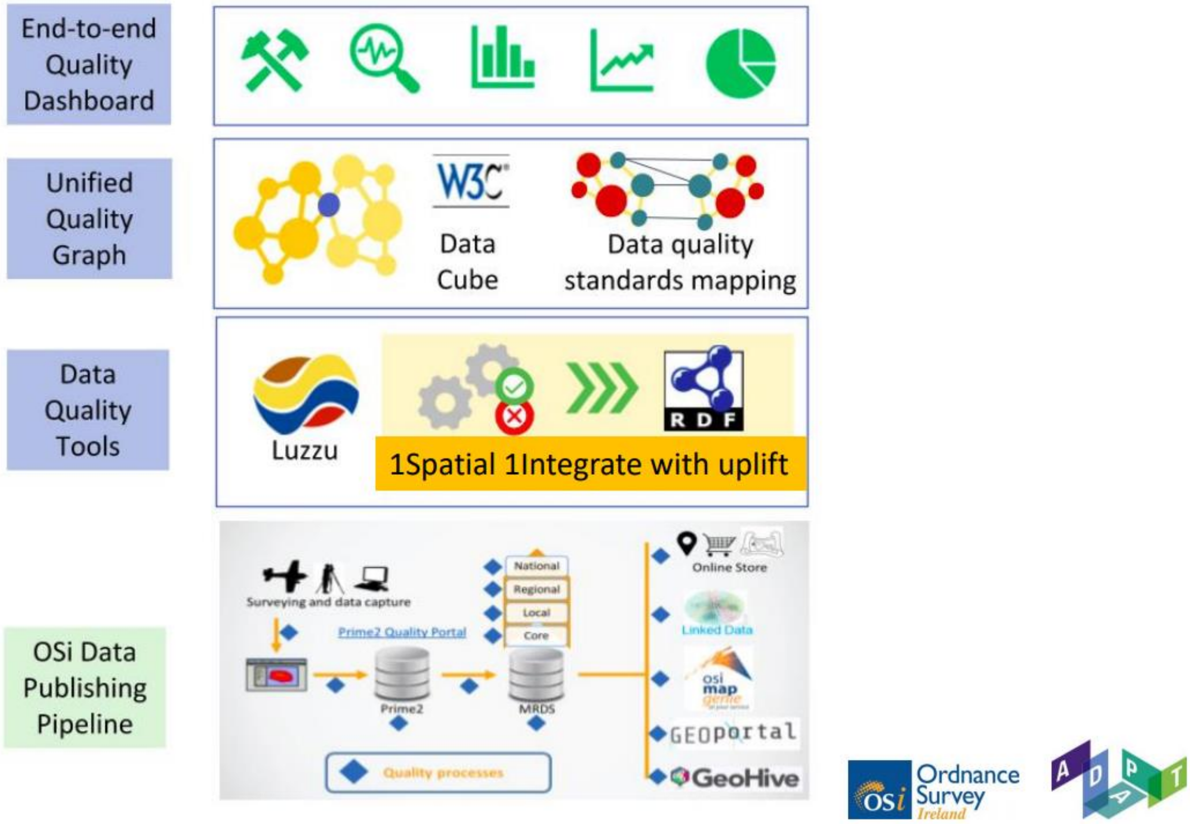
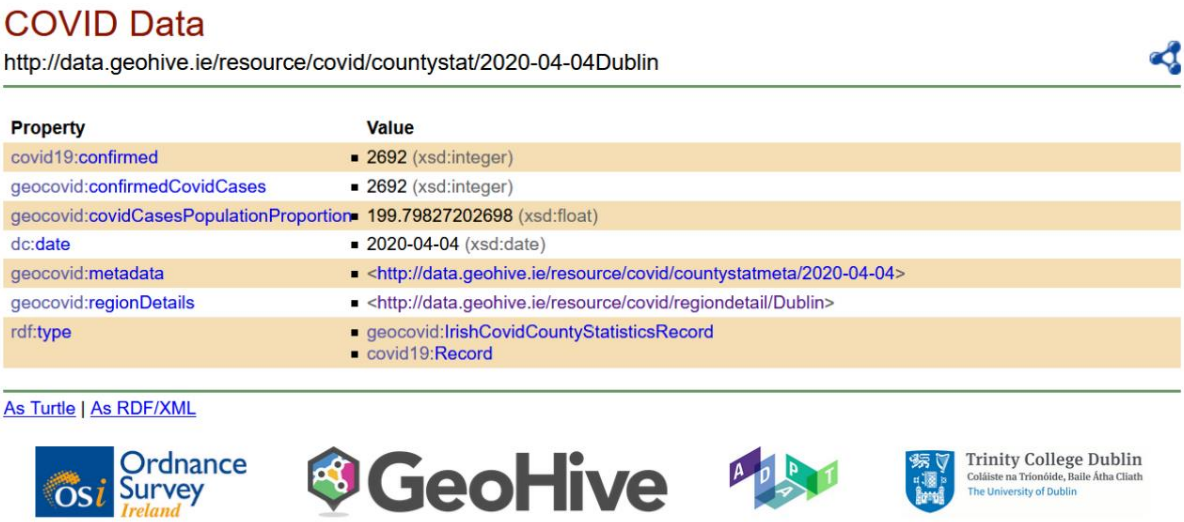


Figure 2: Linked Data for Governance of Geospatial Data



COVID Data

http://data.geohive.ie/resource/covid/countystat/2020-04-04Dublin

Property	Value
covid19:confirmed	2692 (xsd:integer)
geocovid:confirmedCovidCases	2692 (xsd:integer)
geocovid:covidCasesPopulationProportion	199.79827202698 (xsd:float)
dc:date	2020-04-04 (xsd:date)
geocovid:metadata	<http://data.geohive.ie/resource/covid/countystatmeta/2020-04-04>
geocovid:regionDetails	<http://data.geohive.ie/resource/covid/regiondetail/Dublin>
rdf:type	geocovid:IrishCovidCountyStatisticsRecord covid19:Record

As Turtle | As RDF/XML

Figure 3: COVID Data publication © Ordnance Survey Ireland

2.5 swisstopo update (Switzerland)

swisstopo, the Swiss Federal Office of Topography, has been publishing spatial Linked Data since 2017. A first pilot project was launched in 2016 with the aim of building awareness and knowledge on the topic. The project gave good results, so it was decided to put the system into production in March 2017. The Linked Data Service (https://ld.geo.admin.ch) of the Federal Spatial Data Infrastructure (https://www.geoadmin.ch) is serving since then several datasets, such as the Administrative Units (yearly snapshots since 2010 and including generalized geometries), the Public transport stops, the Road traffic accidents and the Statistics of hydropower plants. The system is based on a standard process for the publication of the data according to the Linked Data principles, implying the serialization of the data in RDF, the storage in a triple store and the setup of a Linked Data frontend for URI dereferencing and content negotiation (Figure 4.).

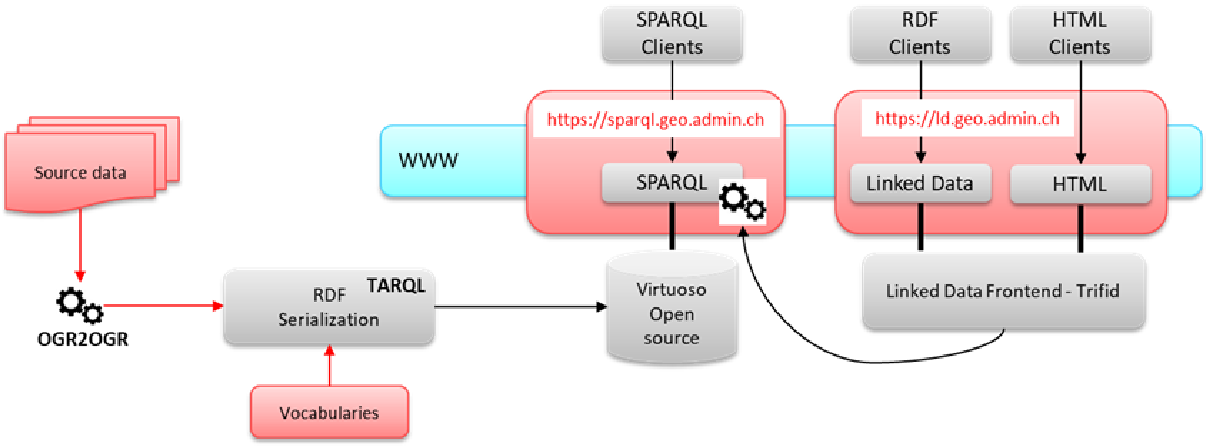


Figure 4: Overview of the publication process

The publication of the Official index of addresses and of the Official index of streets is planned for next year. These are very dynamic datasets, with a daily update frequency, and very voluminous (both contain several million objects), so it was necessary to modify the publication process. Instead of being static, the RDF serialization will take place at runtime thanks to RDF Views configured through R2RML mappings. With this approach, the publication of the data will be synchronized across the various services of the Federal Spatial Data Infrastructure and at the same time redundancy in data storage will be avoided, since all services will be using the same data source (Figure 5).

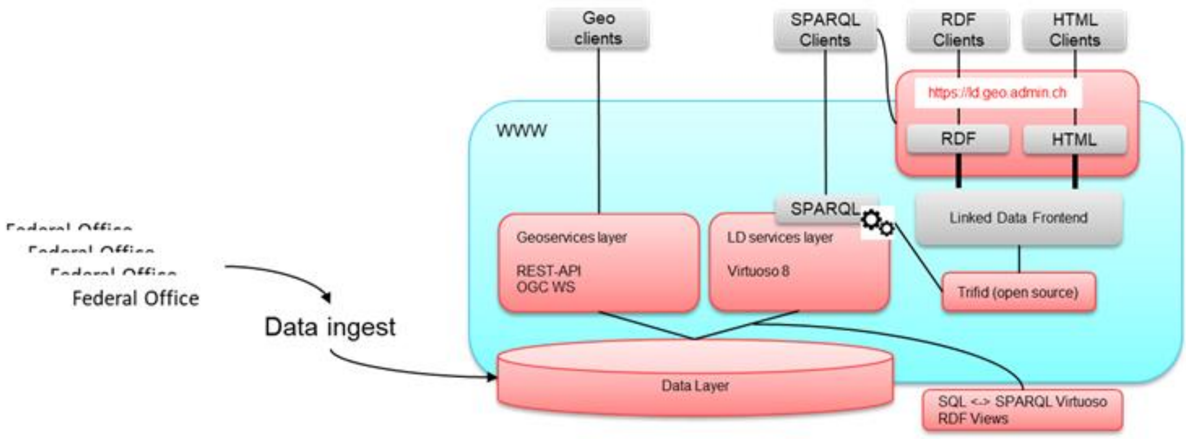


Figure 5: Overview of the optimized data publication process

2.6 Flemish update (Belgium)

The region of Flanders in Belgium runs an interoperability programme ‘[5]Open Standards for Linked Organisations’ (OSLO). The OSLO programme manages the *Knowledge Graph* which includes *spatial information* such as parcels, addresses, buildings, roads, subsoil, mobility, city and road infrastructure (Figure 6).

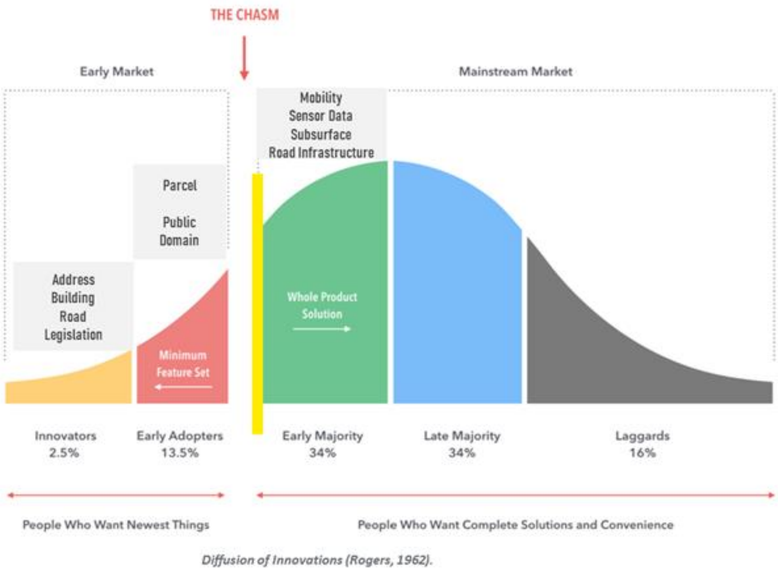


Figure 6: Overview of the evolution of the Flemish Knowledge Graph

The OSLO data specification *process* is aligned with the principles of international standardisation bodies; due process, broad consensus, transparency, balance and openness[6]. The current development activities of OSLO already follow a transparent process: all records of decisions[7] and discussions are publicly accessible[8]. These activities will be formalised and the different process steps adapted to fit the different stakeholders in the specification process, including domain experts, business- and technical analysts. The *method* pursues an implementation of the design principles of linked data[9] as asserted by Tim Berners-Lee in 2006. Existing public sector information systems store data in relational databases and often use Extensible Markup Language[10] schemas to exchange data. These schemas, intended to exchange data, cannot be easily adapted or extended. In the OSLO programme we focus on how RDF, which is an extendable data model, can be adopted in the public sector and how the semantic agreements reached between domain experts, automatically[11] can be transformed into an RDF model preserving the semantic agreements. To allow structured and semi-structured data to be mixed, exposed, and shared across different application[12], it is crucial that the specifications are resolvable on the Web. Therefore, we rewired existing software architectures to a Representational State Transfer (REST) style, which outlines how to construct network-based software applications having the same characteristics as the Web: simplicity, evolvability, and performance. The key innovation lies in combining a bottom-up consensus-based approach with a formal top-down approach which anchors the decisions within a formal government body, using linked data as a blueprint (Figure 7).

---

[5] https://data.vlaanderen.be/standaarden/
[6] https://open-stand.org/about-us/principles/
[7] https://informatievlaanderen.github.io/OSLO/
[8] https://github.com/Informatievlaanderen/OSLO/issues
[9] https://www.w3.org/DesignIssues/LinkedData.html
[10] https://www.w3.org/TR/REC-xml/
[11] https://github.com/Informatievlaanderen/OSLO-EA-to-RDF
[12] https://www.w3.org/2001/sw/wiki/RDF

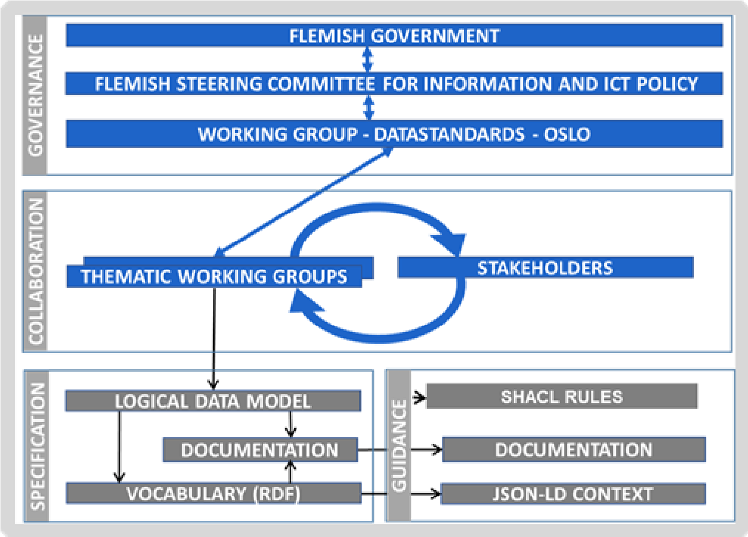


Figure 7: Governance of the Flemish Knowledge Graph

At the time of writing this report the number of ratified information standards by the 'Steering Committee of Flemish Information and ICT-policy' has grown to more than forty[13]. There is a vast adoption of OSLO standards at local and regional government, which *facilitates the integration of spatial data in non-spatial projects*. These projects include the Flemish Citizen Profile[14], Linked Legislation[15] (Local Council Decisions as Linked Open Data) and 'management of the public domain'[16]. Also, we see a broadening to initiatives in the private sector such as Mobility as a Service[17] (MaaS). In 2019 the OSLO process and method was adopted on the Belgian inter-federal level[18] (the federal level, the communities, regions and language areas in Belgium).

To manage the knowledge graph, the Flanders Information Agency and Ghent University have developed an ecosystem of tools, processes and governance for modelling and publishing semantic data standards. The ecosystem relies on Open Source Software and is executed entirely by open continuous integration systems requiring such minimal costs for the organisation. The tooling can be found on GitHub[19]. It allows designing the data standards in a decentral approach, embedded in global governance. The setup works on the basis of separation of concerns: modelling, provenance, a publication with a maximum of automation to facilitate expectations for humans and machines. The core process, usually referred to as OSLO-Toolchain[20], is as follows: the semantical data modelling starts with an annotated UML document. This UML document is converted to human machine-readable and machine-processable artifacts such as HTML pages, JSON-LD context, SHACL files, … The base conversion tools are EA-to-RDF[21] and Spec-Generator[22]. The publication flow is described as a CircleCI configuration[23]. Currently, the OSLO-Toolchain version 2.0 is being in final stage of

13 https://data.vlaanderen.be/standaarden/
14 https://www.vlaanderen.be/uw-overheid/mijn-burgerprofiel
15 https://lokaalbestuur.vlaanderen.be/lokale-besluiten-als-gelinkte-open-data
16 https://data.vlaanderen.be/cms/openbaardomein
17 https://data.vlaanderen.be/standaarden/erkende-standaarden/applicatieprofiel-mobiliteit-trips-en-aanbod/index.html
18 https://github.com/belgif/review/blob/master/Process/201906-ICEG%20-%20process%20and%20method.docx
19 https://github.com/informatievlaanderen/
20 https://github.com/informatievlaanderen/OSLO-Toolchain/
21 https://github.com/Informatievlaanderen/OSLO-EA-to-RDF/tree/json-ld-format
22 https://github.com/Informatievlaanderen/OSLO-SpecificationGenerator/tree/javascript
23 https://github.com/Informatievlaanderen/Data.Vlaanderen.be/blob/test/.circleci/config.yml

roll-out. The reference refers to the corresponding versions for OSLO-Toolchain 2.0. Guidelines and best practices on the modelling the UML are documented here[24]. Other documentation or thoughts can be found in this repository too.

The main driver for OSLO-Toolchain 2.0 was the need for improved decentralisation and versioning support without the loss of overall coherency. Today for each data standard, a new repository can be instantiated[25]. In this sandbox, the editors can apply their way of work to maintain the data standard. By registering publication-ready commits of their sandbox in the central repository[26] a coherent global view is created[27] shows a data standard according to toolchain 2.0 with on the top of the page provenance, transparency and interaction information.

This core is surrounded by tools and other guidelines to facilitate the adoption, such as:

- A URI strategy for persistent identifiers[28] with an implementation on the domain data.vlaanderen.be
- A compliance checking setup: a SHACL validation[29] through a local instance of the ISA testbed validator[30] (English version on JoinUp - EC)
- Technical guidelines for RESTful APIs[31] and supportive tooling for mapping geo-formats[32]
- Process provenance though a registry of data standards[33]
- Tooling for the publication of code lists with persistent identifiers[34]
- A playground to create your own OSLO compliant payloads[35]

This methodology and setup has been applied in real-world projects where the data standards, vocabularies for broad reuse and application profiles for usage in a generic application context, are turned into implementation models using the same toolchain approach. These implementation models are then embedded in deployed software artefacts and APIs. Because the toolchain and methodology is applied in the wide span of data management contexts, and in particular the usage in implementation contexts, creates short feedback loops to the data standards and supportive tools. E.g. we trained the team of Flemish agency for managing the road network[36] (Agentschap Wegen en Verkeer) with a few sessions and cloned the setup for them in[37]. Today the various policy domains are independently integrating their modelling processes to feed this environment. The ecosystem relies on Open Source Software except the Enterprise Architect (Sparx) UML modelling tool. This tool could be replaced by OSS alternatives (Modelio Umbrello UML Modeller, UML Designer, Papyrus, PragmaDev Studio, .etc)

---

24 https://github.com/Informatievlaanderen/OSLO-handleiding/tree/master/Modellering
25 https://github.com/Informatievlaanderen/OSLOthema-template
26 https://github.com/Informatievlaanderen/Data.Vlaanderen.be/blob/test/config/publication.json
27 https://data.vlaanderen.be/doc/applicatieprofiel/mobiliteit-trips-en-aanbod/
28 https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/document/uri-standard-guidelines-flemish-government
29 https://test.data.vlaanderen.be/shacl-validator/
30 https://github.com/Informatievlaanderen/OSLO-Validator-EU
31 https://test.data.vlaanderen.be/doc/richtlijnen/iv-rest-api-richtlijnen/
32 https://github.com/Informatievlaanderen/OSLO-Validator
33 https://data.vlaanderen.be/standaarden/
34 https://github.com/Informatievlaanderen/OSLOthema-lokaleBesluiten/blob/master/.circleci/config.yml
35 https://github.com/informatievlaanderen/OSLO-playground
36 https://wegenenverkeer.data.vlaanderen.be/
37 https://wegenenverkeer.data.vlaanderen.be/

2.7 *Linked Cartography and Maps: the National Geographic Institute of Spain knowledge graph and its semantic web for the public (Spain)*

The project Linked Cartography of The National Geographic Institute of Spain, henceforth IGN Spain, had a main objective: to semantically represent the wide set of geographic resources of the IGN Spain (topographic, regional, word maps at different scales, highest resolution orthophoto, satellite imagery, geodetic and levelling networks and databases, Corine Land Cover maps, LiDAR…) to link them and to publish them in a unified knowledge graph in order to make this rich set of resources and content available to the public in a more useful, friendly, simple, practical, intuitive and interesting way.

The chosen way to do this has been to represent the rich cartographic heritage and aerial photographs archived and published by IGN Spain through a Knowledge Graph that can be interrogated by machines and by people. Geographical resources until then distributed in information silos and non-connected databases, have been grouped, linked, and published as Linked Data, in order to facilitate users the retrieval of information according to their interests and intentions through a metadata search system and intuitive interfaces which always provides contextual information.

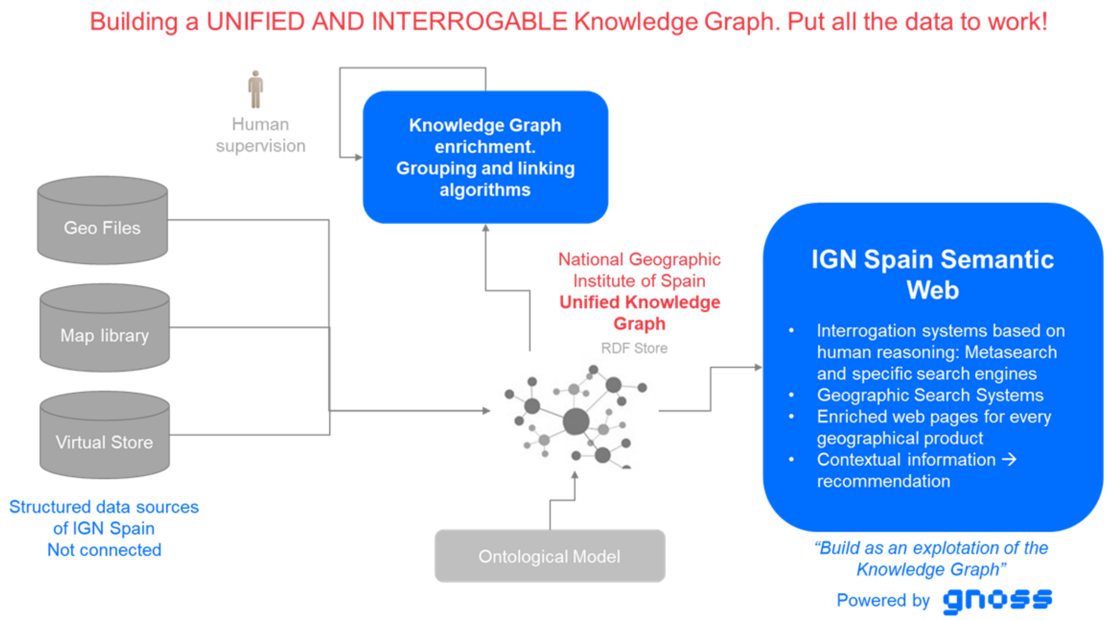


Figure 8: General architectural diagram of the system

The steps for building IGN Spain Knowledge Graph had been:

- To define an ontological model to semantically represent the geographical resources using reference models and vocabularies such as: Regulation nº 1205/2008 as regards metadata based on ISO 19115, Geonames ontology, Good Relations ontology (ecommerce), SKOS and MARC21.
- To build a unified, expressive, extensible and interrogable Knowledge Graph, which groups and links the data stored in three data silos: Geofiles database, Map Library database and Virtual Store database. The result consolidates the IGN Spain data ecosystem as a digital space governed with Semantic Artificial Intelligence techniques.
- To build a semantic web conceived, not only for machines, but specially for people, that exploits the knowledge graph. This approach includes: Iterated interrogation systems which emule human reasoning (Metasearch and facet-based search engines), Geographic Search Systems over maps, Enriched web pages for every geographical product, Contextual information and recommendations.

The result was more than 2M geographical resources semantically represented in a very expressive way, more than 150M triples and a user-centric website to explore and discover this vast volume of information.

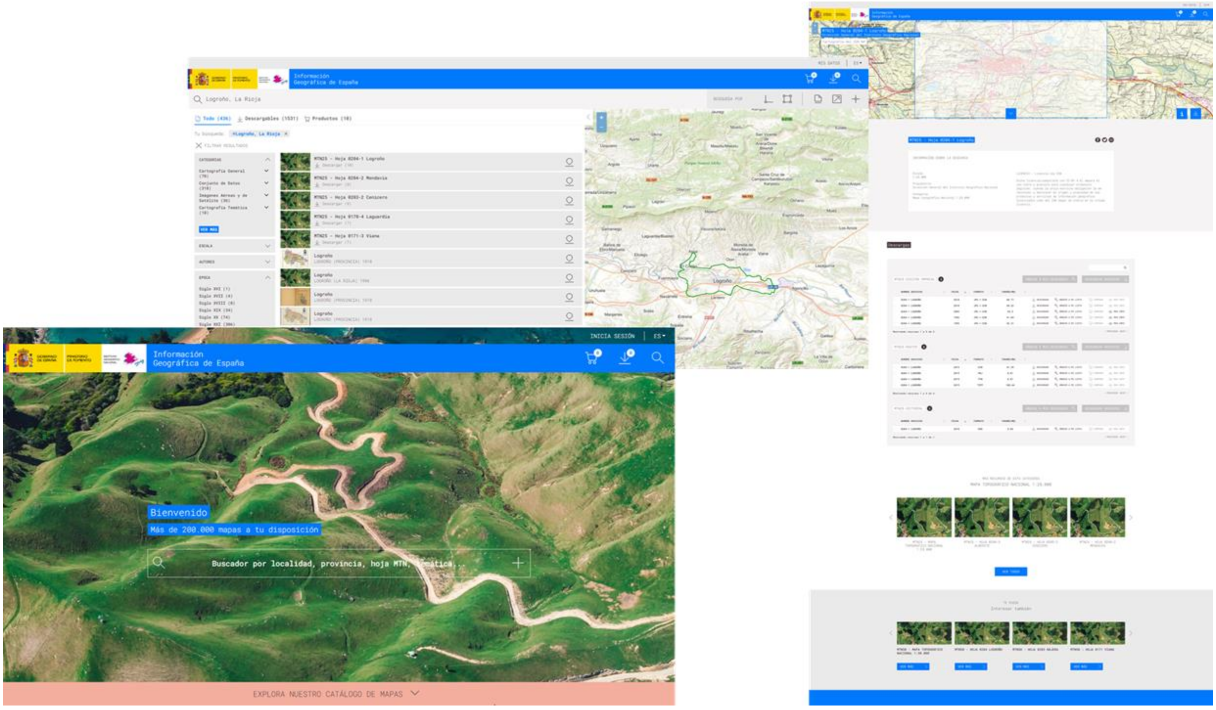


Figure 9: User interfaces of the IGN Spain web semantic application

The main challenges of the project were:

- Defining and building a sort algorithm based on weights to return the most relevant results on the first page of the search engine
- Improving search engine performance: 2 million geographical resources described very expressively implying very complex queries
- Building grouped geographical resources for identifying all distributions (geo files like shapefile, ecw, geotiff, etc) with similar geography and same category or neighbouring categories
- De-normalizing geographical data by grouping geometries to reduce the number of points and simplifying queries
- Converting textual searches into geographical searches thanks to toponym recognition
- Converting textual searches into entity searches thanks to categories recognition

The exciting project of the IGN Spain's Knowledge Graph has been developed by a mixed team made up of IGN staff and professionals from GNOSS company, a technological SME that has been working for more than 15 years in the construction of Knowledge Graphs and the Web of Data (Semantic web) (more info in www.gnoss.com/en), and the technology used for its construction has been the Artificial Intelligence Semantic Framework by GNOSS (www.gnoss.com/en/products/semantic-framework).

3 INTERFACING MORE USERS WITH DATA AND RELATED TECHNOLOGIES

3.1 Self-Service GIS (SSGIS) with Linked Data Tools

Over the course of recent decades, GIS technologies have undergone a rapid transition from networked systems catering to the needs of field experts to web-based applications; leading to the rise of Web GIS and the utilisation of web technologies to make geospatial data, and the visualisation of such, available on the web. The increased number of distributed web mapping applications based on this technology allows better access, interaction and (dynamic) visualisation of geospatial information from heterogeneous sources to a wider range of end users; particularly the non-expert user. Although the transition to web-based GIS has been rapid in pace, arguably a product of the overlap with computer and information science advancements, there does exist a degree of technological disjointness between these fields. The most interesting element of this disjoint is the innovation seen with the Semantic Web, the potential this has within the GIS field and the need for better standardisation if improved integration between web-based applications and geospatial information is to be achieved.

The argument for improved integration of semantic web technologies and web GIS technologies is centred around the fact that the formalisation of geospatial information improves not only the interoperability between different geoinformation sources, but also between spatial and non-spatial data available on the web (Lemmens, 2006). Semantic technologies change the way in which applications interact and access other distributed, heterogeneous sources on the web; improving data processing efficiency and integration of data. In recognising the potential that lies within the development of these technologies, the body of literature relating to geospatial Semantic Web research has focused on improving the exploration, management and linking of explicitly geospatial information with data which contains a spatial element; the goal here being the enrichment of the linked web based on this integration (see Zhang et al., 2015). In line with these efforts, there have been some large-scale adoption initiatives seeking to improve the prevalence of linked spatial data (GeoKnow[38], LinkedGeoData[39]). Despite this, however, the challenge remains as to how best to make use of the web to effectively integrate, manipulate and visualise spatial data in a 'self-service' environment with the goal of bringing this data closer to the non-expert user in particular.

For Kadaster, an integral part of achieving self-service GIS in making use of web technologies is also to embrace the philosophy of openness and connectivity on the web. Indeed, while a self-service GIS necessarily needs to include certain GIS-related functional components, the idea is that it also fully embraces the improved accessibility of geospatial information to all by making use of open standards, interoperability and allowing users access to data 'on the source' in seeking out answers to their geo-related questions. Key to achieving this vision for self-service GIS based on web principles is the standardisation of information on the web according to open standards (W3C and OGC) to improve communication between different (spatial and non-spatial) sources, applications and services on the web. In producing a conceptualisation, the non-expert user group was central to the definition of self-service GIS because of the promise that this type of application has to provide better accessibility for this group in particular based on its ease-of-use premise. It is this group in particular which stands to benefit most from the implementation of this concept.

Exploring the conceptualisation of self-service GIS leads to an enquiry about what features and what functionality is required for the realisation of a self-service GIS. Assessing this necessarily requires an exploration of existing Linked Data tools and what their features are, drawing on those as inspiration for what may be required for a self-service GIS. This analysis is elaborated on in the work of Rowland, Folmer & Beek (2020, under review). The outcomes of this exploration of existing tools highlighted the current limitations of existing linked data tooling options with regards to offering a completely suitable solution to the question of self-service GIS and has lead to a renewed interest in

[38] http://geoknow.eu/

[39] http://linkedgeodata.org/

what tooling Kadaster has already been experimenting with and how these may be adapted to support the challenge of self-service GIS. A realisation of self-service GIS, within singular or plural form, within the context of the Kadaster will certainly make use of the development of such a Knowledge Graph in offering end users the opportunity to access and interact exposed data in a self-service manner, irrespective of expert level and technical background.

3.2 EuroSDR LDGroup initiative : towards a KG about spatial digital assets in Europe

When starting EuroSDR linked data activity in 2015, participants firstly thought that collaboration between Linked data specialists and cartographers would be a natural one. Both communities address the grounding of decisions and knowledge production on (geo)data. And even more than that, designers of the web of data as well just like designers of spatial data infrastructures, and in particular INSPIRE, target that data be produced once and reused several times. Yet, the first exchanges, usually during informal coffee breaks or during round tables, evidenced some mis-communication. Very roughly speaking, on the one hand cartographers would wonder if linked data was yet another buzz word and asking for industrial success stories and not nice to have demonstrators. And on the other hand, linked data specialists would suspect cartographers to be resistant to change and to invent reasons why not adopting LD technology. All in all, some miscommunication was an obstacle to collaboration. Reasons for this miscommunication can be explained by differences in priorities that could lead to conflicts [Bucher et al. 2020]. For example, national mapping agencies are committed to secure geodata supply to support crucial decisions and preserve sovereignty, whereas LD designers are prioritarily committed to secure data reuse and to leverage silos. Besides, while the LD designers adopt the open world assumption promoted on the Web, national mapping agencies have adopted for a long time a closed world assumption in their product design. An important aspect of traditional quality management in geodata is indeed the definition of a contract related to « what to represent and how » between the data producers and the map readers (or the data users). This contract is the product specifications. Quality management is done by documenting the gap between a given dataset and a flawless dataset that would conform to such a contract: the percentage of missing entities or missing attributes, accuracy thresholds for geometric properties and so on. Quality metadata together with product specifications help the user of the dataset evaluate what he/she can infer about reality -what occurs and what does not occur in reality- based on the dataset.

In order to transform conflicts into collaboration, EuroSDR LD group proposes the co-design of an open Knowledge Graph about Geodata in Europe.

Real world use cases of such a KG have been identified:

- use case 1: a scientist/SME has developed an application on some place, for example the French city Toulouse, and would like to adapt it to another place, for example the city Bucharest, or another epoque, for example Toulouse 10 years ago. He/she searches the KG for datasets similar to the datasets he/she already used but with a scope covering Bucharest or covering Toulouse in 2010.
- use case 2: a company searches for data to answer a specific call for tender of the European Environment Agency, like for example producing a specific land cover product.
- use case 3: a water domain expert contributes to defining a European regulation that can be monitored based on existing national data.
- use case 4: an operator of a national geoplatform wants to understand the id policies of different authoritative datasets on a given theme (eg: administrative units) to use them consistently.

The proposed process for the KG creation is firstly to create nodes for some geodigital assets relevant to the use cases as well as “same as” links between them. The creation should be possible for anyone with given authority and knowledge about the digital assets, either using directly EuroSDR LD platform or filling in information on a wizard, and manually create "same as" links (between topographic data products, between DEM with similar resolutions, between ortho-photos, between historical data services, etc). After a feedback with the use case, the nodes will be enriched with properties relevant to use cases, extracted from existing metadata records and the creation of links will also be revised to introduce more specific links and to introduce automatic creation.

3.3 LD Wizard

LD Wizard is a framework developed with the intention of making the transformation of CSV data to Linked Data simpler, through the provision of user-friendly Graphical User Interfaces[40]. It was the result of a joint initiative between Triply B.V., the Dutch Digital Heritage Network (NDE), the Netherlands Cadastre, Land Registry and Mapping Agency as well as the International Institute of Social History (IISH). It is available as an open source product which can be configured to a certain context from the basic "Hello, World!" Wizard. To illustrate how the Wizard works in practice, a transformation of a CSV dataset to Linked Data in three minutes can be found at this link.

4 GeoSPARQL FOCUSED PRESENTATIONS

4.1 GeoSPARQL Benchmark on EuroSDR Sandbox

GeoSPARQL is a standard created by the Open Geospatial Consortium[41] (OGC). The GeoSPARQL introduces a vocabulary for describing complex geospatial information in linked data, and also defines various extensions to the SPARQL query language that allow geospatial information to be queried. As such, the GeoSPARQL standard acts as an important enabler for combining the linked open data (LOD) and geospatial information systems (GIS) domains. The two domains have a lot to gain from each other. Firstly, LOD allows information that is traditionally stored in GIS systems to be integrated with and enriched by other data sources. Secondly, GISs encode detailed spatial relationships and that enrich existing linked data representations.

In its effort to explore novel geospatial technologies that can improve spatial data publication and delivery, the EuroSDR organization wants to explore the viability of contemporary GeoSPARQL implementations. Triply B.V.[42] is a linked data startup that has experience with publishing large GeoSPARQL datasets, e.g. in collaboration with the Dutch Cadastre[43] [Folmer et al. 2020]. Triply has added GeoSPARQL support to the SPARQL services that are exposed in the linked data environment for the Dutch Platform for Linked Data[44] (PLDN). This allows linked data enthusiasts to experiment with GeoSPARQL representations in the linked data they upload, and also allows them to experiment with GeoSPARQL queries that utilize the uploaded geospatial data in innovative ways.

In order to give a decent impression of GeoSPARQL compliance, we need a dataset that contains GeoSPARQL representations and for which GeoSPARQL queries are available. For this the benchmark published in [Huang et al. 2019] is used. This is the most recently published benchmark for GeoSPARQL representations and queries.

The dataset presented in the [Huang et al. 2019] benchmark is published under the EuroSDR in the PLDN linked data environment at https://data.pldn.nl/eurosdr/geosparql-test (Figure 10).

---

[40] See LD Wizard slogan: "From CSV to Linked Data in One Spell".
[41] See https://www.ogc.org
[42] See https://triply.cc
[43] See https://labs.kadaster.nl
[44] See https://data.pldn.nl

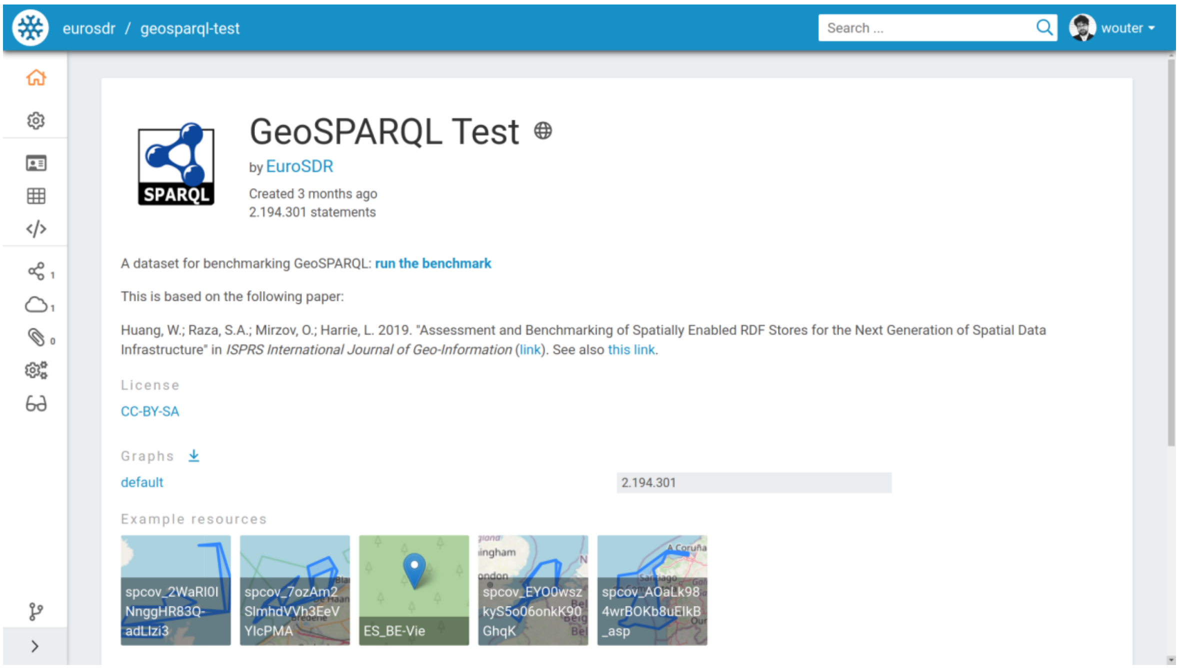


Figure 10: The GeoSPARQL test dataset published under the EuroSDR account in the linked data environment of PLDN (https://data.pldn.nl/eurosdr/geosparql-test).

The GeoSPARQL benchmark dataset is published as open data that everybody can use and play with. The linked data environment provides several functionalities to support its use:

- The dataset can be exported with a simple export button (downwards pointing arrow).
- The dataset can be browsed in a low-level table to see the underlying data patterns, and in a more visually appealing browser that shows geometries and applies quality-printing of various value types (Figure 11).
- Several example resources are included that provided easy access to the linked data browser.
- The dataset is published with good metadata, using both Schema.org[45] (for popular search engines) and DCAT 2[46] (an important community standard for publishing dataset metadata).
- The dataset includes a public GeoSPARQL endpoint that everybody can use. The GeoSPARQL endpoint comes with a feature-rich IDE that supports writing queries and visualizing their results. Specifically, geospatial results can be directly plotted on a 2D or 3D map.

In addition to the data, the queries that are introduced in the benchmark of Huang et al. 2019 are also published under the EuroSDR account of the PLDN linked data environment. These are 27 queries, each of them makes use of a different aspect of the GeoSPARQL standard. Together, these 27 queries provide a good overview of GeoSPARQL compliance.

Since it is arduous to have to run the 27 benchmark queries manually, a data story has been added to the EuroSDR account. The data story (Figure 11) automatically runs the 27 benchmark queries and visualizes their results in an easy to access web page. Visiting the web page in a web browser automatically runs the full benchmark. This makes it easy to observe the live results of the benchmark without having to follow needlessly complicated technological and/or manual steps.

---

[45] See https://schema.org

[46] See https://www.w3.org/TR/vocab-dcat-2

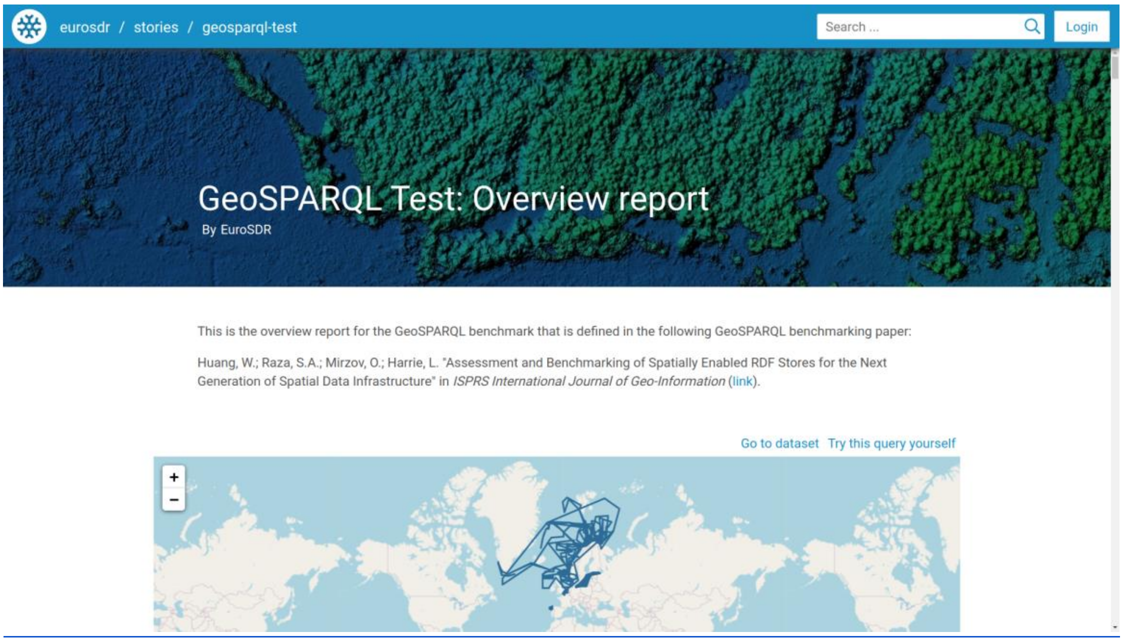


Figure 11: Visiting the GeoSPARQL data story automatically runs the full benchmark. The data story is published under the EuroSDR account in the PLDN linked data environment (https://data.pldn.nl/eurosdr/-/stories/geosparql-test).

In the future, this GeoSPARQL dataset and benchmark should allow linked data and GIS users to more easily test GeoSPARQL by themselves and to further grow the community of GeoSPARQL users. The online publication in the PLDN environment lowers the threshold for starting to experiment with GeoSPARQL. Triply B.V. is using the outcomes of the data story in order to further improve GeoSPARQL compatibility and performance.

4.2 From Buildings to Cities: Multi-scale Geospatial Knowledge Base and Rules

A use case of GeoSPARQL is checking if an existing urban landscape (buildings interconnected with their surroundings) is compliant with urban policies. To address this use case, it is necessary to create a multi-scale spatial knowledge base and connect its layers using geosparql topological relations. Geosparql is an OGC (Open Geospatial Consortium) standard used to represent and query geospatial data on the web, in addition, it supplies vocabulary for asserting topological relations between spatial objects such as simple feature, RCC8 (Region Connection Calculus), and Egenhofer vocabulary.

The multi-scale knowledge base contains the following spatial objects/entities: country, region, department, borough, municipality, district, cadastral, building cadastral, building, river, heat pipeline, sewer pipeline, rain waste, and water waste equipment, etc. The first 6 entities are the French administration levels while the rest are urban components that might exist at a certain administration level.

Geospatial topology helps us define 3 categories of spatial relations:

- Between administration entities for example: which districts are contained within a municipality
- Between administration level and urban components for example: which heat pipeline is contained or crosses over a district
- And finally, between urban components themselves for example: which cadastral parcels contain building cadastral.

With the same method, geosparql can be used to illustrate spatial relation between building elements (storey, zone, space, furniture, wall, slab, etc.) that compose the building scale/entity/concept.

The presentation shows 24 select queries (Qi) that have been applied through two triple stores GraphDB[47] and StarDog[48] to illustrate the spatial relations between above categories. Where Q1 to Q5 stand for spatial relations between administration entities, Q6 to Q12 represent spatial relations between administration entities and urban components, Q13 to Q20 represent spatial relations between urban components, and finally Q21 to Q24 represent spatial relation between building elements.

Query	Parametric	Sub Query	Union
Q1 to Q5 and Q21 to Q23	Y	N	N
Q6 to Q12	Y	N	Y
Q13 to Q20 and Q24	Y	Y	Y

Table 1: table characteristics

As mentioned before, this work used GraphDB and StarDog as triple stores and tools to execute geosparql queries. Authors have noticed the following differences between both tools:

- Syntax: to achieve spatial index in StarDog, you need 'geo:asWKT' as the predicate, while GraphDB do not have this restriction
- Response: GraphDB replies with a Boolean value while StarDog presents the entities that respect the spatial topological requested.
- Geosparql function: StarDog does not support geosparql functions such as buffer, convex hull, union, difference, etc. while GraphDB does.
- Distance calculus: StarDog calculates the distance from polygon centroid while GraphDB uses polygon boundary

Future work aims to:

- Compare geosparql query performance between GraphDB, StarDog and Virtuoso [49]
- Connect building and urban reference system to link their topological relation
- Apply reasoning to building and urban components to link them to different administration levels
- Apply granular computing (indiscernibility) to generate multiple abstractions/views based on the urban rules that need to be checked:
    - Use spatial granularity which will guide us on how spatial relations evolve in case of refinement or reduction of spatial entities
    - Adjust knowledge base new hierarchy, cardinality, concept representation
- Apply logical queries that stand for the urban rules on the different abstractions generated.

[47] https://graphdb.ontotext.com/
[48] https://www.stardog.com/
[49] https://virtuoso.openlinksw.com/

5 DISCUSSION AND PERSPECTIVES

This edition of EuroSDR Spatial Linked Data day evidenced a wider adoption of LD for projects outside the limited field of research. The added value of this technology to more traditional IT used in spatial information infrastructures and web portal is now clearer. One major added value is to support the capacity to integrate data from different sources in applications, thanks to links between these datasets. Applications are national portals powered by knowledge graph interconnecting assets from different authorities, usually thanks to core spatial data. These can be national agencies and statistical surveys for example but also national agencies and Wikidata, the core data framework on the Web. Another added value is to make the authorities and trust framework more transparent to users, which is of crucial importance for some reports like illustrated in Ireland with the COVID19 dashboard.

An important perspective is to empower more and more users with this technology which requires bottom up approach to design vocabularies and more user friendly interface to transform data into LD and query LD.

Another important point is the creation of links between the communities present during the KGiA event, namely spatial linked data specialists, DBpedia and VGI. It was addressed in the Dutch Knowledge Graph and also in the BKG strategy to interconnect its data with Wikidata. An opportunity to do so is the major DBpedia work on referencing existing knowledge graphs.

Last, one pending topic is the scalability to Europe of presented national activities. The EuroSDR LD Group initiative to collaboratively create an open European KG of digital asset can be a step towards scaling to existing platforms.

Acknowledgements

The authors wish to thank reviewers for their constructive comments on the report, Eric Bayers from the Belgian National Geographical Institute, Andreas Hadjiraftis from the Department of Lands and Surveys of the Ministry of Interior in Cyprus, and Jesper Weng Haar from the Dabish Agency for Data Supply and Efficiency. We also gratefully acknowledge the exchanges with the participants and the co-organisers of the Knowledge Graph in Action conference.

References

[Bucher et al. 2020] Bucher, B, Tiainen, E, Ellett von Brasch, T, Janssen, P, Kotzinos, D, et al.. Conciliating Perspectives from Mapping Agencies and Web of Data on Successful European SDIs: Toward a European Geographic Knowledge Graph. ISPRS International Journal of Geo-Information, MDPI, 2020, 9 (2), pp.62. https://www.mdpi.com/2220-9964/9/2/62

[Folmer et al. 2020] Folmer, E, Ronzhin, S, Van Hillegersberg, J, Beek, W, Lemmens, R, Business Rationale for Linked Data at Governments: A Case Study at the Netherlands' Kadaster Data Platform, in IEEE Access, vol. 8, pp. 70822-70835, 2020, doi: 10.1109/ACCESS.2020.2984691.

[Huang et al. 2019] Huang, W.; Raza, S.A.; Mirzov, O.; Harrie, L. Assessment and Benchmarking of Spatially Enabled RDF Stores for the Next Generation of Spatial Data Infrastructure. ISPRS Int. J. Geo-Inf. 2019, 8, 310. https://www.mdpi.com/2220-9964/8/7/310