



Co-financed by the European Union

Connecting Europe Facility

Next Generation of Spatial Data Infrastructure: Lessons from Linked Data implementations across Europe

Stanislav Ronzhin^{1,2}, Erwin Folmer², Roy Mellum³, Thomas Ellett von Brasch³
Eduardo Martin⁴, Emilio Lopez Romero⁴, Samuli Kytö⁵, Eero Hietanen⁵, Pekka
Latvala⁵

¹University of Twente, Faculty of Geo-Information Science and Earth Observation, s.ronzhin@utwente.nl

²Kadaster, erwin.folmer@kadaster.nl

³Kartverket, roy.mellum@kartverket.no, thomas.ellett@kartverket.no

⁴National Centre of Geographic Information, emartin@fomento.es,
elromero@fomento.es

⁵National Land Survey of Finland, samuli.kyto@maanmittauslaitos.fi,
eero.hietanen@nls.fi, pekka.latvala@nls.fi

Abstract

The need for integration of geospatial data across national borders poses questions on how to overcome technical and organizational barriers between national mapping agencies. Existing National Spatial Data Infrastructures (NSDIs) inherited heterogeneous technology stacks and user cultures. Example integration solutions are based on cascading data services on the web using open standards. However, this approach is often cumbersome since it requires substantial efforts aimed at harmonisation of data structures and semantics of the content between NSDIs. In contrast, the Linked data technology as an innovative approach for publishing heterogeneous data sources on the web is able to transcend the traditional confines of separate databases, as well as the confines of separate institutions, keeping existing infrastructures intact. Moreover, exposing national data as Linked Data on the Web makes it a part of the Semantic Web. This allows to shift focus from collection and dissemination of data to meaningful data consumption. Here we present and discuss the results of the Open European Location Services project, a collaboration between the national mapping agencies of Finland, the Netherlands, Norway and Spain which is aimed at demonstrating the capabilities of Linked Data technology in the context of Pan-European geospatial data provision.

Keywords: OpenELS, geospatial linked data, INSPIRE, Spatial Data Infrastructure, SDI

1. INTRODUCTION

Geospatial information has great value for the economy and society alike (e.g. Krek, 2005; Castelein et al., 2010). Therefore, provision of access to European-wide geospatial data is the ambitious goal of the Infrastructure for Spatial Information in the European Community (INSPIRE) (European Commission 2007).

For more than a decade after the INSPIRE directive came into force, the choice of technology to open up geospatial data resources held at the national level was directly related to the open standards developed by the Open Geospatial Consortium (OGC). Based on the eXtensible Mark-up Language technology and the Simple Feature Access model, the OGC standards were understandable and worked well for the community of GIS users.

However, construction of European level Spatial Data Infrastructures (ESDIs) required aggregation and harmonisation of the content coming from national level SDIs. Even though the technical implementation of such aggregation could be done using cascading Web Feature Services, it required resolving the problem of spatial query distribution and cross border fusion of geospatial content (Lehto et al., 2015) which is not trivial.

Another obstacle for increasing reuse of national data at the European level was the lack of adaption of OGC standards outside of the GIS community. Communities with a strong IT background would be more likely to build their applications around RESTful (Representational state transfer) APIs (Application Programming Interface) using JSON-like (JavaScript Object Notation) formats for data exchange.

Therefore, the technical capabilities of the user community together with the way the data is provided from a technical perspective, is crucial for the potential (re)usability of data at any level. Linked Data (LD) technology is proposed as a remedy for solving interoperability problems at technical, organizational and community levels (Janowicz et al. 2012; Ronzhin et al., 2108). The Open European Location Services project, a collaboration between National Mapping, Cadastral and Land Registration Authorities (NMCAs) of Finland, the Netherlands, Norway and Spain was launched to demonstrate the capabilities of Linked Data technology in the context of Pan-European geospatial data provision.

With this paper we attempt to answer several questions related to the use of LD in the context of the INSPIRE initiative, they are as follows:

1. How to solve the problem of initial transformation of data into Linked Data in a pragmatic way without altering national data pipelines and structures?
2. What is the added value of LD-based SDIs for transnational projects?
3. How to demonstrate the potential of LD in the context of European-wide geospatial data access?

2. PROJECT BACKGROUND

There is a long history of efforts to ensure use and reuse of geographical data at the European level. This section briefly discusses the ELS project predecessors, as well as its time line and goals.

2.1. Open European Location Services (Open ELS)

Coming into force on 15th May 2007 (Commission of the European Communities, 2007), the INSPIRE directive *“aims to create a European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment”*. The directive encompasses 34 spatial data themes, grouped into 3 annexes, which can be used to address requirements from environmental applications.

Whilst the INSPIRE directive is a legal and technical framework, the European Location Framework (ELF) (European Location Framework, 2019) project which ran from 2013-2016, was a practical implementation of that directive. Building on the experiences of the European Spatial Data infrastructure Network (ESDIN) (European Commission, 2019a) and European Address Infrastructure (EURADIN) (European Commission, 2019b) projects, ELF was conceived with the aim of *“developing standards, specifications, tools and technical infrastructure to deliver pan-European geospatial content”*.

The Open European Location Services project (Open ELS) builds on key ELF project results. Open ELS is a two-year project co-financed by the European Union's Connecting Europe Facility. It is developing pan-European data services using authoritative geospatial information. In doing so, it aims to improve the availability of geospatial information from the public authorities responsible for mapping, cadastre and land registries. The Open ELS project demonstrates the potential of a single access point for international users of harmonised, pan-European, authoritative geospatial information and services. EuroGeographics, which represents Europe's National Mapping, Cadastral and Land Registration Authorities, is coordinating the two-year initiative co-financed by the European Union's Connecting Europe Facility. The international not-for-profit association is working with partners from member organisations in Finland, Germany, Great Britain, Norway, Poland, Spain, Sweden and The Netherlands to deliver this core

component for wider operational European Location Services. More information is available at the website www.openels.eu.

2.2. Project goals

As part of Open ELS, several sub packages have been defined. In this paper we discuss work package 2.2, which has been defined as:

“Task 2.2 APIs and Linked data. Technical approaches will be developed for making open data easier to use and more flexible through supplementing APIs and demonstrating the “Linked data” concept. ...It is also relevant to support formats more tailored for web-development than the internationally standardised ones, e.g. JSON (GeoJSON) in addition to GML. Geographical names, addresses and administrative unit are good candidates for being exposed as linked data (RDF-format), and OpenELS will demonstrate how this can be done. In addition, persistent identifiers will be tested and published as RDFs and SPARQLs.”

The work package defined five goals or milestones by which the work could be measured, these are:

1. Define common ontologies for the AU and GN themes
2. Create linked data in RDF format and make it available through a SPARQL endpoint.
3. Upload URI's to the geolocator application
4. Develop a data story to showcase the LOD
5. Write a report detailing the development and results of the work package.

3. RELATED WORK

Topics discussed in this paper feature several important concepts that need to be defined. This section introduces them following (roughly) their chronological order of development.

3.1. Linked Open Data

The LD initiative (Berners-Lee et al., 2001) promotes the use of semantic standards for representing and publishing information on the Web at data level. This implies that each data attribute is individually recognisable, retrievable, and combinable into aggregate statistics (W3C, 2014a). This can be achieved by encoding information using the Resource Description Framework (RDF) (W3C, 2014b). This standard is based on mature technologies: the graph data model (Silberschatz et al., 1996) and the Hypertext Transfer Protocol (HTTP) (Fielding et al., 1999). The former allows instances and concepts, represented by nodes, to be related to one another by relationships, represented by arcs between the nodes. Through HTTP Universal Resource Identifiers (URIs) these data elements (nodes

and arcs) become globally accessible, referenceable (Bizer et al., 2009a) and queryable by the means of the SPARQL query language (Perez et al., 2006; W3C, 2013).

3.2. Geospatial Linked Open Data

Many data suppliers (Folmer & Beek 2017), especially ones that publish official government data, such as NMCAs, are diving into the world of linked data as they see potential benefits for their authoritative data.

Ordnance Survey, the national mapping agency of Great Britain, was one of the first big governmental organisations to pioneer exposing public geospatial data on the Web as Linked Data in 2008 (Goodwin et al., 2008). Even though this was a state-of-the-art development at that time, it relied on unstandardised means for representing data semantics and as a result lacked (re)usability. The Ordnance Survey Ireland evaluated the experience of their British colleagues and used standard vocabularies to publish boundary data of administrative division at various level and to capture the evolution of administrative boundaries (Debruyne et al., 2017). Another prominent example is the work of the National Geographic Institute of Spain (IGN-E) (de Leon et al., 2010) where they combined data coming from two governmental institutions and published it as a coherent Linked Data dataset. Important work was done by Battle & Kolas (2011) that introduced a geo extension to SPARQL, namely GeoSPARQL.

Since many SDIs use the GML standard for data encoding, automated generation of LD from GML sources was proposed and developed by van den Brink et al. (2014). Comprehensive guidelines based on the best practices for publishing, retrieving, and using spatial data on the web were developed by van den Brink et al. (2018).

3.3. Bridging Linked Data and SDI

Schade et al. (2010) identified two common scenarios of Linked Data usage within SDIs and conclude that only minor changes to current SDI standards are required for implementation. Links can be either embedded at the service level or at the feature level by means of content negotiation. In the latter, WFS may offer its data in classical GML, in RDF, or in HTML depending on the accessing client. However, this approach does not allow to reason over the data using Semantic Web reasoners. This drawback can be solved by specifying transparent and bidirectional proxies that allow users of both infrastructures to share data and services (Janowicz et al., 2010). Portele et al. (2016) designed and implemented an intermediate layer using proxies that made data and metadata from the OGC web services available on the Web of data. Availability of Linked Data resources via WFS protocol was tackled in the work of Jones et al., (2014) by developing the

LOD4WFS Adapter, a software that transforms WFS requests into SPARQL queries on the fly.

3.4. (Re)use case potential

Since Linked Data is a relatively new technology for most users, many of them are unaware of the potential that can be unlocked. To bridge the gap between (a) the vast but implicit potential that a Linked Dataset encapsulates, and (b) the specific and often more explicit use cases a prototypical user may have in mind, a concept of Data Stories was introduced and developed by Folmer and Beek (2017). A Data Story allows a specific use case to be explained to a potential user through a sequence of data examples, that are connected by an overarching story. To be as generic as possible, the data examples that compose a Data Story are visualizations of SPARQL result sets. This ensures that the components of a Data Story are declarative (how the data is obtained is encoded in the SPARQL query), reproducible (the query is recomputed when the Data Story is generated), and modifiable (advanced users can click a button to open the SPARQL query view, where the query can be altered and rerun).

Following the Link Data Visualisation Model developed by Brunetti et al. (2013), Data Stories provide functionality to apply appropriate visualisation techniques depending on the results of queries, e.g. tables, diagrams/charts, pivot tables, widget galleries, or geo-spatial maps. As such, many different types of information, e.g., geographical and statistical information, can be combined to tell an engaging story with data.

4. STATE OF LINKED DATA IN MEMBER ORGANISATIONS

Four NMCAs co-operated in the development related to the OpenELS project goals introduced in Section 2.2. Table 1 provides names, origin and web address of the participant organisations.

Table 1: The member participants of the OpenELS project

Name of organisation	Explanation	Web site
Instituto Geográfico Nacional (IGN)	National Geographic Institute of Spain	http://www.ign.es
Kadaster	Netherlands Cadastre, Land Registry and Mapping agency	https://www.kadaster.com
Kartverket	Norwegian Mapping Authority	https://www.kartverket.no
Maanmittauslaitos (MML)	National Land Survey of Finland (NLS)	https://www.maanmittauslaitos.fi

4.1. Norway

The Norwegian mapping authority currently has 1 national dataset available as Linked Open Data (LOD) through a production environment, Administrative Units. This is currently being utilised by the national geoportal (<https://geonorge.no>) for indirect geocoding of dataset coverage within metadata records. It is also planned that the national data catalogue will begin to use the data and associated services within 2019. Future plans include delivery of the national place names dataset as LOD. This dataset has more potential to be used nationally, given that it would support many more use cases, not least as a stable and authoritative gazetteer.

Currently, the technical architecture includes two database servers for the Virtuoso graph stores (test and production); one application server for the rest api, ontology store and content negotiation; and one proxy server. The ontology creation and data conversion (from a relational database to RDF) has been completed on local machines.

While LOD is currently in use within the organisation, the technology itself is still undergoing an acceptance process to evaluate if it is an area that will be actively supported in the future. Questions within the process surround the resources required, the use cases it solves and other technologies that it might replace.

4.2. Finland

There have been several prototypes to publish Linked Open Data in National Land Survey of Finland. Geographical Names, Buildings and Administrative Units have been published as demo services. Also, URI services to provide spatial objects from dereferenceable HTTP URIs as linked data and human readable HTML pages have been implemented. There has been a prototype of a URI service which uses WFS2.0 as a data source as well as a URI service which uses a SPARQL endpoint as a data source. In 2015, a national recommendation for unique identifiers of the geographic information has been published which defined the structure model of the URI identifiers and the practices of redirection. In addition, NLS-FI is currently implementing WFS3.0 compatible services, for example on the GN theme.

4.3. The Netherlands

The Netherlands have a production Linked Data platform (<https://www.pdok.nl/>) in use for publishing several actual and maintained linked datasets such as Key Register Addresses and Buildings, Key Register Topography and Cadastral Parcels. This production environment consists of a complete stack of tools covering all processes from creating, to publishing, linked data. The Research and Development environment (<https://data.labs.pdok.nl/>) provides experimental tooling which is different from the production environment.

4.4. Spain

The National Geographic Information Centre has developed a platform for the dereferencing of entities through the generation of Persistent Identifiers (PIDs) from its INSPIRE WFS services. In addition, a SPARQL endpoint has been created with the aim of generating and storing the RDF triples in Turtle format for these same themes, as well as its link to other Linked Open Data, such as DBpedia and Geonames. This endpoint is based on Parliament, a W3C solution, and allows SPARQL queries.

In addition, efforts are being made to promote the use of LOD among the different Geographic Information producing organisations in Spain. Currently, a working subgroup has been created to document an inventory of platforms or projects based on LOD, as well as a series of use cases and a guide to generating this kind of information.

5. APPROACH

The idea of Linked Data is to achieve interoperability based on the standards and principles that have been defined in the linked data context, but without standardisation of framework/tools/vendors, etc (Ronzhin et al., 2018). To demonstrate this property of LD, we chose not to have one common approach for preparing the LD in this project. Instead we focused on achieving the common result of publishing Linked Data (RDF) according to common ontologies, URI structures and individual SPARQL endpoints. Some key choices made within these areas were:

1. *Regarding generation of RDF.* We used the RDF Guidelines from the EU's Joint Research Center: <https://github.com/inspire-eu-rdf/inspire-rdf-guidelines>
2. *Regarding the use of ontologies.* We used the INSPIRE ontologies for Administrative Units and Geographical Names. <https://github.com/inspire-eu-rdf/inspire-rdf-vocabularies>
3. *Regarding URI strategy.* We used INSPIRE IDs to coin URIs. However, we use valid country specific domain names, which implies that all URIs are dereferencable. Because of the experimental nature of the project we did not put constraints on the domain part of the URIs since there was no expectation to create persistent URIs. For the same reason, we published one version of the data without supporting history, versioning or updates.
4. *Regarding SPARQL endpoints:* A SPARQL conformant endpoint, preferably including Geosparql support (Battle & Kolas, 2011).
5. *Regarding geometries:* We use WGS 84 coordinates for geometries because this fits best with our government status and potential client capabilities.
6. *Regarding linking to external resources.* To have an improved showcase we decided to include links to Dbpedia from each of the datasets.

By taking this approach it became an interesting opportunity to see if, and how, easy interoperability can be achieved in practice.

To prove this, we used common Open Source software components to build a demonstration web page. The page was constructed around YASGUI (Folmer, et al., 2018), an integrated SPARQL editor and result set visualizer developed by Triply (<https://triplify.cc>). In collaboration with the Kadaster Data platform, YASGUI was extended to support GeoSPARQL. With this extended support it is possible to query for geospatial relationships, return them in a standard-compliant result set, and automatically display them on a 2D Leaflet map. YASGUI was also extended with the support of national background maps of Spain, Norway and Finland. Ideally, we would have added the openELS Basemap to provide a pan-European map, but this was not possible due to licensing issues. The demonstration page features a Data Story presenting a number of use cases to explain the value of LD to a potential data (re)user.

6. RESULTS

All four participating countries generated Linked Data versions of the two INSPIRE datasets, namely Administrative Units and Geographical Names. These data can be queried via national SPARQL endpoints. The results can be visualised on the country's own official base maps. To store relevant queries, and to demonstrate the results, we have created a Data Story. This section provides further detail on the methods and tools used to generate and publish the data. Quantitative characteristics of the resulted datasets are given followed by the description of the GeoLocator service. The data story is presented at the end of the section.

6.1. Methods and tools

Table 2 provides an overview of the methods and tools used for generation of the LD. In all cases, it was a one-time transformation of the content of the source data providing Administrative Units and Geographical Names. However, as can be seen from Table 1, all participants used different tools and frameworks for transformation. The teams from Finland and Spain chose Python as the main software environment to develop custom scripts for data transformation. In contrast, the Netherlands and Norway opted for standalone desktop software packages to perform semantic mapping of the original source content.

Diversity of the possible software solutions that can be used for data publication was also demonstrated by the fact that all participating countries used different backend applications including Jena Fuseki, GraphDB, Virtuoso Open Source, and Parliament Open Source.

Table 2: Methods and tools used for data transformation and publishing

Countries	Transformation		Publishing	
	Method	Tools	Endpoint (software)	Dereferencing (software)
Finland	One-time-transformation	Custom made Python script (RDFLib) from PostGIS to Jena TBA	Jena Fuseki	Python Django framework with RDFLib
The Netherlands	One-time-transformation	OpenRefine with RDF extension	GraphDB	SNORQL
Norway	One-time-transformation	Karma - open source data integration desktop application with API.	Virtuoso open source	Python flask
Spain	One-time-transformation	Custom made Python script: parsing GML features	Parliament open source	PIDMS

6.2. Administrative Data

Table 3 provides information about the generated data sets containing Administrative Units. As can be seen from the table, Finland, Norway and the Netherlands have a similar number of units and comparable sizes of the datasets. In the case of Spain, the resulting dataset was almost twenty times larger because of the high total number of Administrative Units.

Table 3: Overview of datasets containing administrative units

Country	Admin levels and their names	Number of admin units (total)	Total number of triples	Size in MB (for rdf/xml serialisation)	Example URI
Finland	Level 1: Country (1) Level 2: Regional State Administrative Areas (7) Level 3: Regions (19) Level 4: Municipalities (311)	338	8752	8	http://paikkatiedot.fi/so/openels/au/AdministrativeUnit/091

The Netherlands	Level 1: Country (1) Level 2: Provinces (12) Level 3: Municipalities (388)	401	5211	30	http://data.labs.pdok.nl/dataset/openels/au/AdministrativeUnit/NL.BRK.AU.0363
Norway	Level 1: Country (1) Level 2: Provinces (18) Level 3: Municipalities (422)	441	5748	30	https://data.geonorge.no/openels/au/AdministrativeUnit/id/0
Spain	Level 1: Country (1) Level 2: Regions (19) Level 3: Provinces (52) Level 4: Municipalities (8218)	8290	220435	148	http://datos.idee.es/recursos/openels/au/AdministrativeUnit/1760959

6.3. Geographical Names

The datasets containing Geographical Names were almost one thousand times larger than for Administrative Units. Table 4 summarises the quantitative characteristics of the results of the transformation.

Table 4: Overview of datasets containing geographical names

Country	Total number of toponyms	Total number of triples	Size in MB (for rdf/xml serialisation)	Example URI
Finland	~1500000	8501970	1990	http://paikkatiedot.fi/so/openels/gn/GeographicalName/40472569
The Netherlands	~1000000	10000000	2100	http://data.labs.pdok.nl/dataset/openels/gn/NamedPlace/NL.TOP10NL.GN.130456316
Norway	949615	15817780	2763	https://data.geonorge.no/openels/gn/NamedPlace/654818
Spain	1132743	18491679	2400	http://datos.idee.es/recursos/openels/gn/NamedPlace/176097

6.4. GeoLocator

A URI lookup service was created by utilizing a database model of the ELF GeoLocator service. This is a gazetteer service containing place names data from various European countries. The URI lookup service was implemented first by creating a new GeoLocator database instance. The instance was populated with the data from the Geographical Names and Administrative Units themes of the four participating countries. After that, the URIs that were created in the project for the Geographical Names and Administrative Units objects were added to the database.

The URI lookup service was implemented as a Java servlet that executes fuzzy name searches to find near matches to the inputted keywords. The service response contains the URIs together with some general information on the spatial objects, such as the coordinates and the feature type classifications.

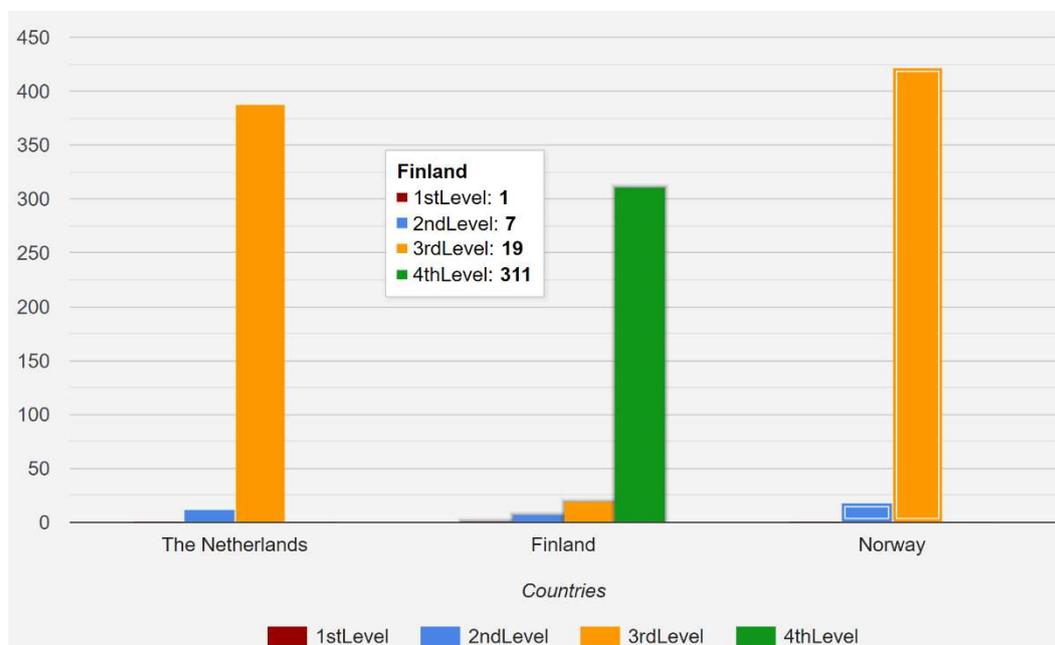
6.5. Data story

A Data Story allows the original data publishers to emphasize the potential use cases that they envision for their dataset. This includes their ability to highlight interesting aspects of the dataset itself, e.g., interesting objects and/or interesting relationships between objects, as well as interesting ways in which the data can be combined with other Linked Data sources (e.g., DBpedia). A Data Story can be thought of as an 'advertisement tool' for data. In this section we present the data story available at <https://data.labs.pdok.nl/stories/OpenELS/>.

6.5.1. Semantic interoperability

Often, it is difficult to compare structures of administrative division between countries because they use language-specific names for levels of administrative division. Linked Data brings semantic interoperability by means of ontologies. With the help of a common ontology (in this case the INSPIRE Administrative Units ontology), it's possible to model national systems of administrative units and draw comparisons between them, avoiding misinterpretation caused by language specific notions. Figure 1 illustrates the comparison of the structure of administrative systems in the Netherlands, Spain, Norway and Finland.

Figure 1: Number of administrative units per administrative level

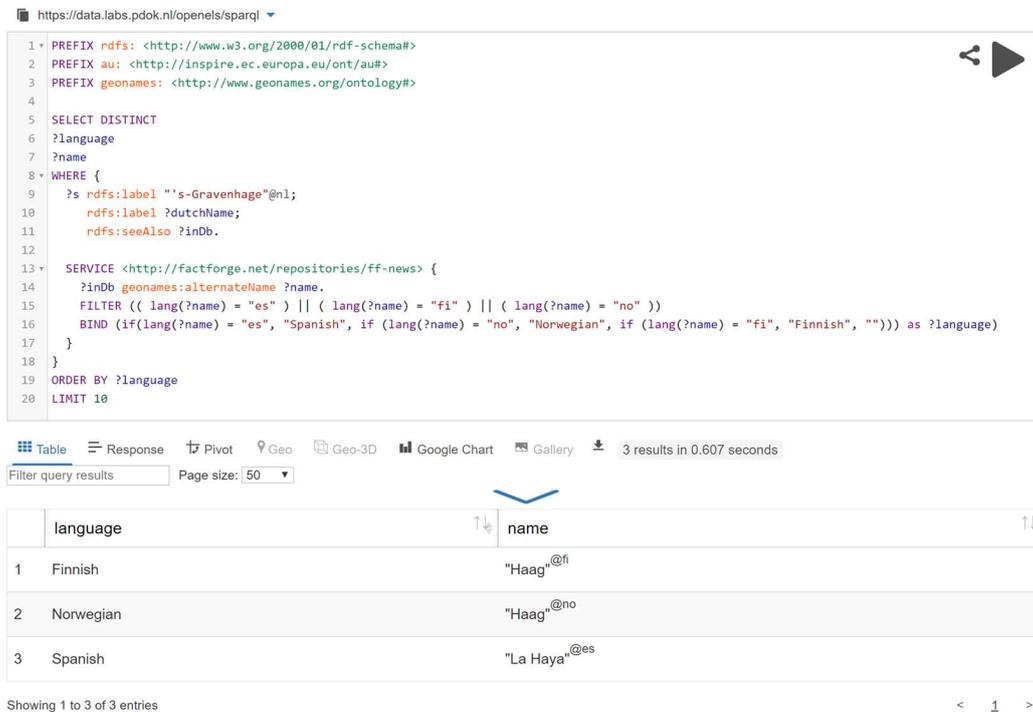


Source: Authors

6.5.2. Translation of geographical names

In the example above, we used the power of linked data to map national administrative systems to generic levels. This was done at the level of concepts or ontological level (like T-box in descriptive logic), but Linked Data allows for even more - interrelating data at the instance level between data sets. For example, if two different data sets contain information about the same object (e.g information about the same municipality) we can link them and enrich one description of the municipality with attributes coming from the other data set. The example provided in Figure 2 uses links between administrative units data maintained by the NMCAs and corresponding objects in the DBpedia data base. In this case DBpedia serves as a 'linking node', providing access to other resources and information. Therefore, it is possible as an example, to traverse those links to retrieve the name of an administrative unit in another language. Figure 2 demonstrates a query retrieving the spelling of what the Dutch call "s-Gravenhage" in the other languages of the project (Finnish, Norwegian and Spanish). The results of the query are presented as a table.

Figure 2: Translation of geographical names.



Source: Authors

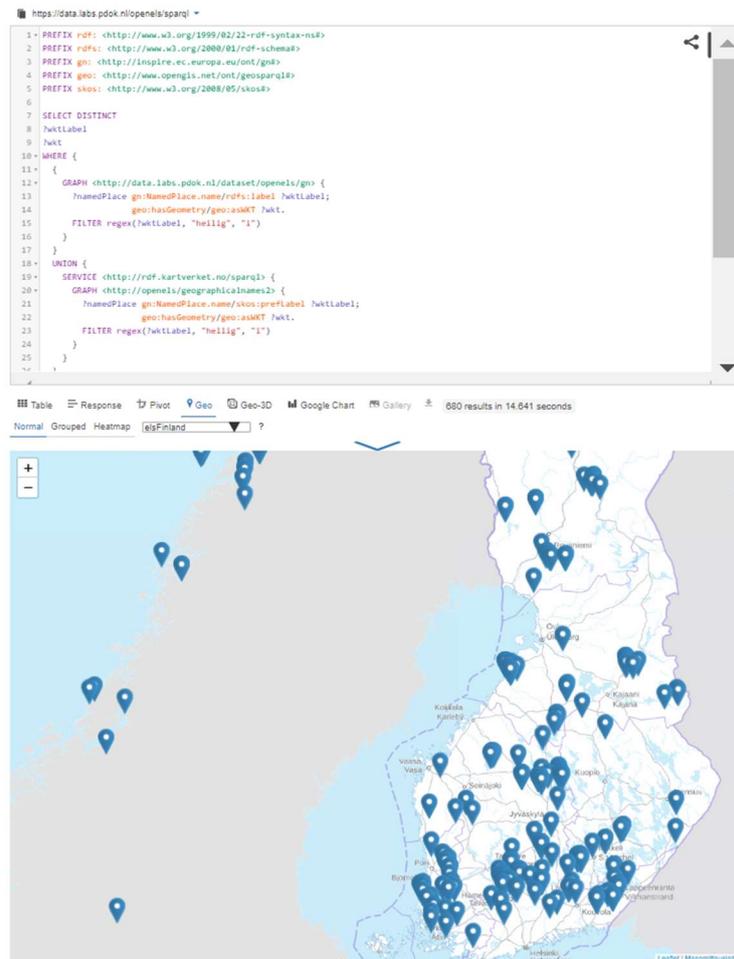
6.5.3. Holy Geographical Names

The OpenELS project has published more than 40 million place names (toponyms) and locations as Linked Data. Standardised concepts for describing the meaning of place names enables semantic interoperability between national data sets. Seamless access to such rich data allows for conducting interesting research. For example, linguists could potentially use this data to analyse the spatial distribution of common toponym roots. The following example (Figure 3) shows locations of places that have “holy” as part of the name. Obviously, the root “holy” is spelled differently in different languages (see Table 5). The Linked Data technology makes it possible to formulate a single query that can interrogate the national endpoints in their native languages.

Table 5: Translation of “holy” into languages of project participants

Language	Spelling
English	Holy
Dutch	Heilige
Norwegian	Hellig
Finnish	Pyhä

Figure 3: Visualisation of the results for the query retrieving toponyms that feature “holy” as part of the name.



Source: Authors

7. LESSONS LEARNED

7.1. INSPIRE-based ontology as starting point

The INSPIRE Data Specification guidelines, defined various GML application schemas for different INSPIRE themes. In general, the data model harmonization process usually requires that some compromises are made regarding the harmonized data content. The data models usually differ at the national level and all data elements cannot be included in the harmonized model.

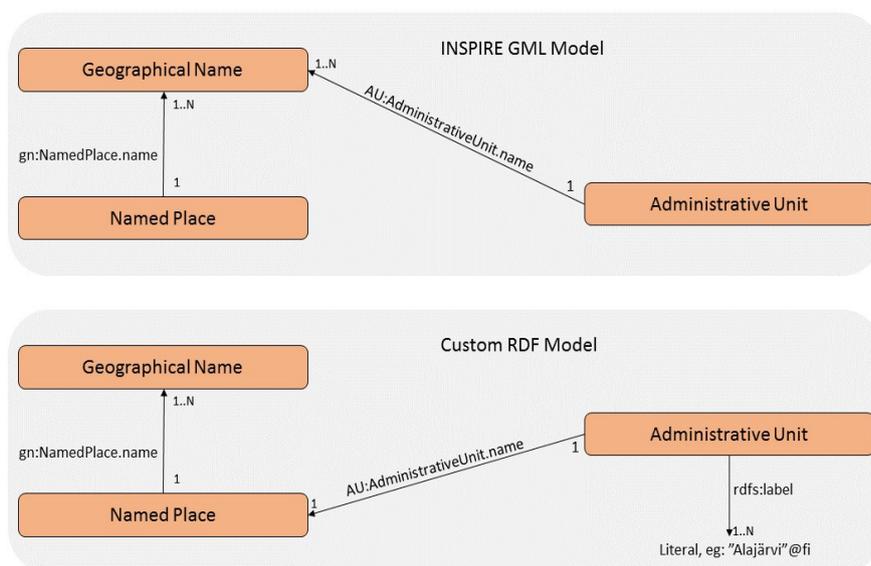
The INSPIRE ontologies that were used in this study contained the same properties as the INSPIRE GML schemas. Transforming data from the harmonized INSPIRE data models to linked data was feasible. However, the linked data representation does not need to be restricted to the INSPIRE content, and can be easily extended with additional links or national content.

The INSPIRE schemas allow the creation of application domain extensions, but that requires the use of custom schema definitions that can cause interoperability problems in applications. A Linked data-based approach gives more flexibility through extended data content, as additional information can be more easily added. The open world assumption in Linked Data enables the data content to be enriched also for rare use cases. However, further work is required for defining ontologies and vocabularies required in the geospatial domain. Also, the use of proper validation methods is still required for achieving harmonized datasets at the European level.

7.2. Flexible data model

The disadvantage of the INSPIRE GML schemas is that they often contain complex data structures that are rather difficult to understand. In comparison, the linked data representation contains only ordinary links between resources instead of complex elements. The data models of the linked data can also be adjusted more easily. This was tested in the project by making minor changes to the semantics of the data model. For example, as shown in Figure 4, the INSPIRE property "name" for AdministrativeUnit refers to a Geographical name object. In the Linked Data representation, an Administrative Unit can be linked straight to the corresponding Named Place object, which consists of multiple representations (e.g. different languages). In addition, we can use *rdfs:label* and *skos:prefLabel* (to link to the actual names) which both support different languages and are not "domain level", but rather more general definitions.

Figure 4: INSPIRE GML Model and Modified RDF Model



Source: Authors

7.3. URI strategy

In this study we discussed URI patterns, but did not see any benefits through harmonising the URIs. Regardless, we agreed on common pattern principles, such as using Theme name and Featuretype according to the INSPIRE recommendations and a LocalId, for the URI path. Different countries use their own domain names and we decided to keep those.

7.4. Geographic coordinates and projections

Using only the geographic WGS84 coordinate reference system (CRS84) and OGC's Well Known Text (WKT) syntax in this study avoided any unnecessary technical issues regarding support of other CRSs. This was possible since we didn't have any use cases which would have needed multiple CRS support. Named places have point geometry and Administrative units have (multi)polygon geometry. Different countries are able to deliver different scales for the geometries. The default geometry for Administrative Units was decided to be the most accurate one available. Geometries are implemented as blank nodes or as object type resources with an own URI. This allows for attaching metadata to the geometries which is important if we want to provide different representations e.g. different scales. Making the geometry an object type makes it more flexible to store and query the

data. One can for example save the geometries with their own URIs to a separate RDF graph to increase performance. OGC's GeoSPARQL ontology defines some terms to provide geometry and metadata information. Still this is not sufficient e.g. as there is no term for geometric scale.

7.5. Data from the source

The data used in this study were generated using a one-time transformation approach. The major limitation of this approach is its inability to update the datasets easily and regularly. In some countries (e.g. Netherlands) updates of INSPIRE data are lagging behind the updates of national data sets. Therefore, creating yet another clone of those data worsens the problem. The future work should include creating data transformation pipelines that would support automated incremental data updates and versioning.

8. CONCLUSION AND FUTURE WORK

The OpenELS project experimented with the provision of geospatial Linked Data in the context of European-wide data dissemination. The four participating countries successfully transformed two source datasets according to INSPIRE themes, namely Administrative Units and Geographical Names, into Linked data and deployed the necessary infrastructure for data publishing. It was concluded, based on the experience, that LD technology is able to provide solutions for solving technical interoperability. However, additional work is needed to achieve better semantic interoperability, e.g. by developing higher-level ontologies (i.e., ontologies that would assist in alignment of national administrative levels).

As was shown in Section 6.1 there was not a unified single solution for initial data transformation. A range of software packages were used, demonstrating flexibility of the approach towards data transformation. Availability of the INSPIRE ontologies was considered as an important prerequisite that helped with the initial transformation. However, with the aim of avoiding versioning problems, it is necessary to develop better synchronization between INSPIRE/OpenELS URIs and national URIs and their versioning.

Flexibility of the LD data model discussed in Section 7.2 adds significant simplification to the process of creating Pan-European data services. Local peculiarities of existing national data schemes can be nullified within the resulting data services at almost zero costs.

The Data Story approach was used to showcase the advantages of LD in an interoperable way. The GeoLocator service was built to demonstrate the advantages of injecting LD URIs into a semantically poor data service.

At the time, all tasks done have been related to Administrative Units and Geographical Names. Considering the current conclusions, the same process carried out should be extended to the rest of the INSPIRE themes. Having a collection of different themes, the next step should be the relationships between them. As a proposal, a first experience of that could be linking Administrative Units with Geographical Names, or Hydrography with Transport Networks.

At other levels, it is possible to explore how to link the external data from agencies that work with non-spatial information, such as EUROSTAT or national statistical agencies and institutes. In order to have a stable community that consumes linked data and generates new applications based on it, the make-up of future working groups is proposed to be at the national level

9. ACKNOWLEDGEMENT



**Co-financed by the Connecting Europe
Facility of the European Union**

10. REFERENCES

- Battle, R., & Kolas, D. (2011). Geosparql: enabling a geospatial semantic web. *Semantic Web Journal*, 3(4), 355-370.
- Berners-Lee, T. (2001) Linked data: Design issues, at <http://www.w3.org/DesignIssues/LinkedData.html>, [accessed 27 March 2019].
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. <https://doi.org/10.4018/jswis.2009081901>
- Brunetti, J. M., Auer, S., García, R., Klímek, J., & Nečaský, M. (2013). Formal linked data visualization model. In *Proceedings of International Conference on Information Integration and Web-based Applications & Services* (p. 309). ACM.
- Castelein, W., Bregt, A., & Pluijmers, Y. (2010). The economic value of the Dutch geo-information sector. *International Journal of Spatial Data Infrastructures Research*, 5, 58-76.
- Commission of the European Communities (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007. Establishing an

Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal of the European Union, L108: 1-14.

- Debruyne, C., Meehan, A., Clinton, É., McNerney, L., Nautiyal, A., Lavin, P., & O'Sullivan, D. (2017). Ireland's Authoritative Geospatial Linked Data. In International Semantic Web Conference (pp. 66-74). Springer, Cham.
- De León, A., Saquicela, V., Vilches, L. M., Villazón-Terrazas, B., Priyatna, F., & Corcho, O. (2010). Geographical linked data: a Spanish use case. In Proceedings of the 6th International Conference on Semantic Systems (p. 36). ACM.
- EuroGeographics (2019). About Us, at <https://eurogeographics.org/about-us/>, [accessed 27 March 2019]
- European Location Framework (2019). Project Home Page, at <http://www.elfproject.eu>, [accessed 27 March 2019].
- European Commission (2019a). European Spatial Data Infrastructure Network (ESDIN), at <https://inspire.ec.europa.eu/SDICS/esdin>, [accessed 27 March 2019]
- European Commission (2019b). EUROpean Addresses Infrastructure (EURADIN), at <https://inspire.ec.europa.eu/SDICS/euradin>, [accessed 27 March 2019].
- Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., & Berners-Lee, T. (1999). Hypertext transfer protocol--HTTP/1.1 (No. RFC 2616). doi:10.17487/RFC2616, at <https://tools.ietf.org/html/rfc2616>, [accessed 27 March 2019].
- Folmer, E., & Beek, W. (2017) Kadaster Data Platform - Overview Architecture. Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings: Vol. 17, Article 23. Available at: <http://scholarworks.umass.edu/foss4g/vol17/iss1/23>
- Folmer, E., Beek, W., & Rietveld, L. (2018). Linked Data Viewing as part of the Spatial Data Platform of the Future. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42(4), 49-52.
- Folmer, E., Beek, W., Rietveld, L., Ronzhin, S., Geerling, R., & den Haan, D. (2019). Enhancing the Usefulness of Open Governmental Data with Linked Data Viewing Techniques. In Proceedings of the 52nd Hawaii International Conference on System Sciences. <http://hdl.handle.net/10125/59728>
- Goodwin, J., Dolbear, C., & Hart, G. (2008). Geographical linked data: The administrative geography of Great Britain on the semantic web. *Transactions in GIS*, 12, 19-30.

- Janowicz, K., Schade, S., Bröring, A., Keßler, C., Maué, P., & Stasch, C. (2010). Semantic enablement for spatial data infrastructures. *Transactions in GIS*, 14(2), 111-129.
- Janowicz, K., Scheider, S., Pehle, T., & Hart, G. (2012). Geospatial Semantics and Linked Spatiotemporal Data – Past, Present, and Future. *Semantic Web*, 3(4), 321–332. <https://doi.org/10.3233/SW-2012-0077>
- Jones, J., Kuhn, W., Keßler, C., & Scheider, S. (2014). Making the web of data available via web feature services. In *Connecting a digital Europe through location and place* (pp. 341-361). Springer, Cham.
- Krek, A. (2005). Geographic information as an economic good. In *GIS for sustainable development* (pp. 105-124). CRC Press.
- Lehto, L., Latvala, P., & Kähkönen, J. (2015). Cascading Geospatial Services for integration of authoritative national datasets-CASE: European Location Framework. In *Proceedings of the 27th International Cartographic Conference*, CD-ROM.
- Open European Location Services (2019). Project home page, at <https://openels.eu/>, [accessed 27 March 2019].
- Pérez, J., Arenas, M., & Gutierrez, C. (2006). Semantics and Complexity of SPARQL. In *International semantic web conference* (pp. 30-43). Springer, Berlin, Heidelberg.
- Portele, C., van Genuchten, P., Verhelst, L., & Zahnen, A. (2016). Spatial Data on the Web using the current SDI. Report of the research results in the Geonovum testbed “Spatial Data on the Web” (topic 4), at <http://geo4web-testbed.github.io/topic4/#h.82m7wlpX09r6> [accessed 27 March 2019].
- Ronzhin, S., Folmer, E., & Lemmens, R. (2018). Technological Aspects of (Linked) Open Data. In B. van Loenen, G. Vancauwenberghe, & J. Cromptvoets (Eds.), *Open Data Exposed* (pp. 173–193). The Hague: T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-261-3_9
- Silberschatz, A., Korth, H. F., & Sudarshan, S. (1996). Data models. *ACM Computing Surveys*, 28(1), 105-108.
- Schade, S., Granell, C., & Diaz, L. (2010). Augmenting SDI with linked data. In *Workshop On Linked Spatiotemporal Data*, in conjunction with the 6th International Conference on Geographic Information Science (GIScience 2010). Zurich, 14th September.(forthcoming 2010).
- Van den Brink, L., Janssen, P., Quak, W., & Stoter, J. (2014). Linking spatial data: semi-automated conversion of geo-information models and GML data to RDF. *International Journal of Spatial Data Infrastructures Research*, 9(2014), 59-85.

Van den Brink, L., Barnaghi, P., Tandy, J., Ateazing, G., Atkinson, R., Cochrane, B., Fathy, Y., García Castro, R., Haller, A., Harth, A. & Janowicz, K. (2018). Best practices for publishing, retrieving, and using spatial data on the web. Semantic Web, (Preprint), 1-20.

World Wide Web Consortium (2013). SPARQL 1.1 Overview, at <https://www.w3.org/TR/sparql11-overview/>, [accessed 27 March 2019].

World Wide Web Consortium (2014a). Best Practices for Publishing Linked Data, at <https://www.w3.org/TR/ld-bp/>

World Wide Web Consortium (2014b). RDF 1.1 concepts and abstract syntax, at <https://www.w3.org/TR/rdf11-concepts/>, [accessed 27 March 2019].