

Report on first EuroSDR Linked Data seminar

France, Paris; 2015, April 10th

Authors:

Rijsdijk, M., Kadaster, The Netherlands.

Bucher, B., IGN France, France

Arjen Santema, Kadaster, The Netherlands

Abstract

The first EuroSDR Linked Data seminar was organized by the Dutch Kadaster and IGN France in Paris in April 9th-10th 2015. Objectives of this working session were 1) to get more insight in the potentials of LOD technologies for NMCAs and other organisations in charge of producing and distributing authoritative data, 2) to get more insight in the new needs for location framework brought up by these somewhat disruptive technologies, and if these are these true needs (or just buzz) and 3) to define future actions among participants and especially, for EuroSDR, to propose a future role for EuroSDR.

It was organized in two days. First day was joined with another event organized in Paris by the Datalift association to gather French participants having different experiences of Linked Data benefices and stakes. Datalift association promotes the technology of linked data and especially the road map of the datalift open source software that can be used to publish data as linked data. It was organized as a series of conferences and small tutorials around the datalift platform.

The second day (10th) was a working session with a more limited number of participants and organized after EuroSDR guidelines (small presentations and break up sessions). These participants included members of other organisations than NMCAs because Linked Data is about linking different contents together, not only among one theme across several countries, but also among one common space across several themes and points of views.

A conclusion was that there is a need to connect different European initiatives related to authoritative data and linked data. Technical objectives for a common road map were identified. Besides, organizational stakes were also identified: to organize a dialogue between different profiles to let a fruitful new community emerge among Linked data believers and people attached to GI science unsolved issues, to reach more users thanks to Linked data technologies (what users).

1 Feed back from the Datalift day (april 9th)

This day was an opportunity to attend presentations from various early adopters of Linked data technologies. It was in French but some tutorials that were partly in English to support the understanding of EuroSDR participants. The program and presentations can be downloaded from http://datalift.org/?page_id=153.

Fabien Gandon from INRIA proposed a broad overview of latest trends on the Web, including provenance metadata, access management, licensing issues. He also described the functionalities provided by the datalift platform. Several tutorials and a presentation of the software architecture datalift were organized to facilitate participants adoption and familiarization with this open source software. Bernard Vatant presented an initiative that gained more and more interest from the Web community, the LOV (Linked Open Vocabularies).

Among thematic presentations, Nathalie Abadie (IGN-France, COGIT lab) analysed the current situation of geographical 'content' distribution on the web and the need for a proper vocabulary to publish geographical reference data and publish explicit connections of thematic content with this referential. (<http://datalift.org/downloads/3%20-%20Voculaires%20g%C3%A9ographiques%20-%20Nathalie%20Abadie.pdf>). Franck Cotton from the statistical institute presented the publication of statistical data on the web of data. Thomas Saint Aubin from La Dila (Public organization in charge of distributing legal information) presented a more organizational perspective on e-administration and linked data and La Dila approach to foster innovation in this domain, mainly by connecting actors together around challenges.

2 Feed back from the working session (april 10th)

2.1 Presentations

The working session gathered fifteen participants; some with a professional career in geodata research, from Slovenia, Netherlands, France and Switzerland and others with a professional career in authoritative data research and innovation from French Statistical Institute, National Library, Geological Survey, Water management information system.

Five presentations were given and gave insight in dutch and french nation-wide linked-data initiatives. They can be downloaded from EuroSDR website and some are detailed hereafter.

The presentation from the Dutch Kadaster showed the way to find unstructured information on the internet and how Unique Resource Identifiers (URI's) that are Unique Resource Names (URN's) are used in the web infrastructure. The Dutch Cadastre was taken as an example and the Dutch URI strategy was presented.

- <domain> for example <http://bag.kadaster.nl> (where BAG is the Dutch abbreviation of the Key Register on Addresses and Buildings)
- <type> id, doc or def. Things have an id, for example a specific building (in Dutch: 'pand'). If one types the identification (type: id) uri of a specific building, one gets 'some useful information', namely the documentation (uri type: doc) about that building. Def uri's consist the information about the onlology of the data, the data model.
- <concept> defines the type of thing, for example 'pand' (building)
- <reference> contains the reference to an instance of a concept, for example the building with identificationnumber [0200100000001088](http://bag.kadaster.nl/id/pand/0200100000001088)

For example, <http://bag.kadaster.nl/id/pand/0200100000001088> URL returns the page that contains the description of the building with this identifier. The methodology followed for this project was to start by structuring all the knowlegde about a registration in a SKOS vocabulary. Next step was to build a

data ontology in which the concepts in this taxonomy are realised (<http://bag/kadaster.nl/def#>). Concepts can be linked to sources such as the law. Data have metadata such as the way they are collected, trustability, etc.

The presentation from BRGM, the French office for geological information, insisted on the two aspects of standards that can be legally binding and that need to be community agreed. Specific work on controlled vocabulary for mineral resource in Inspire was presented and their usage in the project OneGeology Europe to browse data. BRGM may use URIs for different resources : Document / reports, Datasets / an object within a dataset (even an observation), A controlled vocabulary / an entry within a controlled vocabulary, Metadata, feature catalogue, etc... In their domain, labels are bad identifiers because ambiguous so they have to use identifiers at one point. Besides, they use compatible InspireId and DOIs. DOIs are useful because organisations can change names and they support the connexion between former and new InspireId. An important stake is to be more precise in the semantics of SameAs links and to be able to describe also what is different and why two representations are needed instead of merging them.

The presentation from BNF, the French national library, described the usage of Linked Data technology at the national library and its impact on the visibility of BNF web site from search engine and its ranking. Geographical entities appears twice in BNF vocabularies since they may be locators (place) or features of interest (subject). Different queries using the graph of resources were presented.

The water management system (ONEMA, France) from France presented their URI scheme and <http://id.eaufrance.fr/>. First URIs have been defined for the referentials (<http://id.eaufrance.fr/apt/1234>) for dictionaries. (<http://id.eaufrance.fr/ddd/par/2.3>) and for scenarios which describes processes (<http://id.eaufrance.fr/scn/quesu/3.0>).

IGN-France presented its experience with Linked data, mainly acquired during the collaborative research project Datalift funded by the French agency for research (ANR) which resulted into the Datalift platform and association. IGN defined vocabularies, URI schemes and set up a prototype semantic server to publish some of its resources :

- Spatial coordinate reference system (a vocabulary that is currently reused by the W3C, and its usage to describe specific systems like <http://data.ign.fr/id/ignf/crs/RGF93G>)
- Features used in spatial indirect referencing like administrative entities (<http://data.ign.fr/id/geofla/departement/51>), reused for instance by the french statistical institute to align their data sets with the topography
- Schemas : <http://data.ign.fr/def/geofla.html>, [../geometry.html](http://data.ign.fr/def/geometry.html), [../topo.html](http://data.ign.fr/def/topo.html).

IGN also developed methods to convert geometries into RDF and to relate geometries and any feature (hasGeometry property). We studied some scenarios of mediation between a geoserver and a semantic server in order to avoid duplicating the data stores. Interconnection of thematic data with reference spatial data can be done either by defining geometries for the thematic data or by defining explicit relations between thematic data and spatial data besides sameAs links. NearOff links were proposed by IGN.

2.2 Road map definition (break-out sessions and their conclusion)

Second part of the working session was reserved for break-out sessions. Objective was to bring focus in European research for linked data and identify items for a road map and possible next steps : what are the key issues. A list of research topics was generated and these research topics were then clustered.

Scientific and technical research issues in linked data research were defined:

- Geostandards; which should be in scope; Search to the standards of European Uniform Resource Identifiers;

- How to manage time and space. How to deal with versioning?
- How to display information to people (more graphical, more textual) ?
- Autodata matching and how to organise this;
- Ranking which is an important issue (ranking vocabularies by relevance to a context, ranking resources answering a sparql query, ..);
- Educating users, and allowing feed back from users of URIs to publishers of URIs.

Some *more organisational* issues were also defined:

- How to organize the dialogue between those who are believers and those who are not believers and also between promoters of different complex technologies. How to share best practices european wide together when the specific mixed community does not exist (meaning people did not reach a common ground);
- what is the position of Inspire wrt Linked Data;
- How to access the economical benefits brought by INSPIRE;
- Should there be still new protocols and how to anticipate it?
- How should this become more user-driven?

At this moment there are different national linked data platforms and also different groups (like NMAs, Geological surveys, libraries, ..) studying linked data. Each of them have their own community, experiments and scope. From this perspective, the conclusion at the end of the first break-out session was that it's not realistic to define one european and united research plan. However, countries and organisations could still learn from each other. EuroSDR could contribute to connect different linked-data initiatives in Europe with the objective to learn from each other, using the spatial aspect as an integration use case (the base map where different linked data can connect).

2.3 Identification of use cases (break-out sessions and conclusion)

In the afternoon session, participants decided it should be a great idea to answer the defined research questions by "living" user-cases. Some of these user cases with connection to Linked data technology were defined:

- Tourism; wanting information of a place where you are. How do you reach this (technology, actors) ?
- App-generator; supporting the fast development of application by developers who don't have much knowledge of spatial information complexity (GML specification for instance).
- Participation of citizens; how to enrich authoritative data with different information sources on the web?
- History of families; how to link different datasets?
- Crossborder projects; important research question is how to deal with semantic problems?

The cases crossborderprojects, app-building and Insight in the history of family were chosen as the "most successful" potentials. Crossborder relates to Europe and INSPIRE. App-building relates to connecting 'communities' (people familiar with linked data and people familiar with gi techno) and to users. Families connects thematics with places and support addressing time and space, as well as legal documentations (birth certificates and so on). Idea is to develop these cases further by one or more institutes or organisations related to the EuroSDR network. Results of the breakout sessions could be handled over to them who become supervisor of these linked-data usecases.

4. Practical conclusions

There is a need to connect different European initiatives related to authoritative data and linked data. The above list of suitable users-cases could be use to set up possible sandboxes or sessions to do this.

EuroSDR has to organise a one-time 2-days Linked Data Seminar at the end of 2015/ begin of

2016. If EuroSDR wants to stay a frontrunner in data research; it should stimulate to connect the different European initiatives in linked data. There is no European initiative to bring best practices together till so far;