Position paper Linked Data EuroSDR

2014, December 29.

Authors:

Bucher, B., IGN France, France

Hansen, T. Brigsted, Danish Geodeta Agency, Danmark

Rijsdijk, M., Kadaster, The Netherlands.

1. Introduction

Since a few years the topic "linked (open) data" (LOD) shows up at EuroSDR meetings. Therefore it was decided to invite a key-note speaker, i.e. Erwin Folmer from the Netherlands. He was there during the 124th EuroSDR delegates meeting (May 14-16, 2014, Apeldoorn, the Netherlands) to explain what one can do with LOD. It went all expected developments like why it would be relevant for EuroSDR members and what EuroSDR activities should be started on this topic to support the use of LOD by NMCAs. In a break out session, it was decided to prepare this positioning paper to find further answers to these questions.

The objective of the positioning paper is to get more insight in a) the domain of LOD, b) potentials of LOD technologies for NMCAs, incl. in terms of new usages for authoritative location framework brought up by these technologies, and c) ongoing initiatives that operate on the border of geo-ICT and LOD. From these insights the position paper should conclude on relations to EuroSDR-interests and in possible EuroSDR activities to be stared that could be beneficial for NMCAs that want to introduce LOD within their organisations.

This paper will give EuroSDR and its delegates directions how to get involved in this topic next years to bring it to the world of NMCAs.

First, this positioningpaper gives an explanation about linked (open) data is. Then, it reports about ongoing activities on Linked (Open) Data in The Netherlands, Denmark and France followed by some critical reflections. At the end, the position paper will draw conclusions and recommendations for possible follow up actions. The position paper will be presented and actions will be discussed during the 125th EuroSDR delegates meeting in October 2014, Leuven, Belgium.

2. Definitions

The themes "linked data" and "open data" are not the same. Wikipedia says "The goal of linked data is to give more meaning to information by defining the specific contexts (website, Geonovum, 2014). For example, linked data gives content to words like the city of "Paris" or the attribute "Geodesy". To give it this content, the real meaning of the word has to be defined. After this more content will lead to other related words and definitions and descriptions with relations to this content.

This will lead to much more meaning and effect in the improvement of digital data search.

The technical foundations of linked data are :

- Using persistent http URIs (Uniform Resource Identifiers) to refer to objects of interest, setting up services so that these http URI actually returns meaningful description;

- Publishing links between data sets, typically through reusing URIs from one dataset for another or through the publication of statements “sameAs” between URIs;

- Using rdf format to structure data (a graph structure);

- Publishing, interlinking and reusing data structures (called “vocabularies”).

The meaning of Open Data is defined as much as “open for anyone”, which means the data must be available under a free licence. Data is financed by public funding and should be distributed to every citizen and such called open standards will be used.

So Linked Open Data could be seen as the sum of two, which will mean open free public data, which is optimally linked to eachother and could be found by using the real meaning of the word during search-operations. Importantly, the W3C ranking scheme for LOD sets up as high priority the 92nd

publication on the web with open licence and open format before the usage of Linked Data principles.

From a critical view, Linked Open Data will stimulate the public society. At this moment, public society has lots of data, but datasets are not linked to each other.

From another critical view, Linked Open Data is promoted by the W3C consortium and not the OGC consortium. Even though W3C shares some objectives with OGC they have a different business background and made different technological choices like choosing graph structure to store and query information instead of layers inherited from our cartographical history. Their objectives differ slightly in the usage. OGC aims at interoperability between components belonging to the GI world. W3C aims at the interoperability between connected components, which is a larger domain. Hence, these techno are adopted by a much wider community.

3. Ongoing Linked Data initiatives in some member states

Situations at some delegates EuroSDR

France

In France, some public data are delivered using URI and RDF : data from the french statistical office (INSEE, http://rdf.insee.fr/demo/index.html), data from the national library (BNF, http://data.bnf.fr/en/). The cultural domain is very interested in linked data technologies.

In the domain of geographical data, several actors have started studying this new technology and is comparing this with existing technologies. At IGN-F, following up previous work (3 phds) on the formalisation of geographical data semantics at COGIT lab, researchers have studied relevant “vocabularies” for a spatial reference framework for Linked Data in coöperation with computer scientists. IGN-F is also contributed to an open source platform to publish data as linked data (datalift) and experimented it to set up data.ign.fr, publish vocabularies and a sparql end point. Ongoing work at IGN-F relates to the interconnection of data on the semantic web : establishing connections with data published by the statistical office and the national library, and providing methods to enhance the interconnection of data based on semantics.

Besides, a cluster gathering indutrials from the information sector (GFII, http://www.gfii.fr/en/) has set up an initiative to learn more about semantic web technologies and has selected the theme of water for this.

Denmark

In Denmark the public administration is mid 2014 in the early implementation phase of a Common Data Distribution Platform for Basic Data, which has Linked Data as an option. Nevertheless, the vision on how to use, operate and develop the technology is in its infant stage. The first very

precautious steps will be to decide a strategy for the use of it and to raise awareness of the possibilities for the data owners and start the hard work to harmonise and specify the rules for the common use of technology. There is in spite of the current situation no doubt that Linked Data and supporting technologies will be used sooner or later. What will be important to evaluate in the upcoming strategy is the business needs for the technology. Later, the way to implement and maintain the linking mechanism will be needed and last: to get all necessary data entities being wrapped in RDF and set up the operating environment to support dereferencing of links as well as the maintenance of data and systems.

The Netherlands

In the Netherlands researches focuses at linking public data of the key registers. Topics and documents like juridical restrictions, communication-documents and sites will be linked to each other. There will be zoomed in the datasets which are content in the key "standard" registers like addresses, buildings, topography and land registration. Geonovum, the national Dutch institute of geo-information and spatial research cooperates in a pilot in cooperation with different partners. The Dutch Cadastre is involved in it as a holder of the key register Buildings and Addresses. The pilot has an objective to give partners more knowledge and a first impression on how public data could be linked to each other. The Pilot-group has defined seven themes in which the partners will be working on and have linked these themes to seven cases. These cases contains research to user interfaces, semantics and how to relate apps for example. Each case will be ended with a short symposium; during these sessions results are presented to stakeholders and anyone who has interest in it.

4. Reflections

Collaborative research projects addresses the acquisition, management and delivery of spatial data and services while international workshops and courses, in collaboration with related organisations, addresses key issues in a timely and focussed manner. Vision of EuroSDR is "To be the European research platform for National Mapping and Cadastral Agencies, Academic Institutes, the Private Sector, Industry and User Groups. The platform focuses on issues related to the implementation of technology developments with respect to optimising the provision, collection, processing, storage, maintenance, visualisation, dissemination and use of geographic information".

From this perspective, there is a direct link between Linked Data research and the field of interest of EuroSDR. Next to topics like the development of 3D, optimising generalisation and improvement of datastorage en indices, more knowledge and activities in the field of linked data could support EuroSDR to keep the frontrunner in spatial data research.

At this moment, NMCA's are changing. People expect quality of NMCA's spatial data, these data are really trustable. Next to that, users of geodata become more and more self supportive. Finally, data of NMCA's will become more and more integrated in a larger field of other datasets outer the geoscene.

Lots of NMCA's anticipate directly or indirectly to these perspectives. For example in the Netherlands; research will be done in how to link different datasets topography with different scales. NMCA's are becoming more customer driven than ever, they make products that society really wants; more and more geoproducts have been made together with potential customers themselves. In the old days, NMCA's collected data and static products were made. In the actual days, datasets are collected and stored. As a result of questions from the society (themes like safety, environment, health care eg) static and dynamic data will be collected and transferred into information that society really need and can re-use. The actual way of indexing data is not good enough, new innovative ways have to be developed. One of them is linked data; it could be a solution for better searching and collecting data.

So, sharing located data on the web is an important stake; the market beyond is supposed to be huge and it will facilitate the management of our territory at different scales with data more varied and possibly more up to date. Next to the web, datasets like INSPIRE or ELF are also contributed to this stake. INSPIRE and ELF are both focused on data provided by legal entities. These consists of local technology based on a federated schema and on standards promoted by the GI software industry and data providers (OGC, ISO).

From these perspectives, important questions that NMCA's face today are :

- How to provide persistent and dereferenceable uniform resource identifier for which objects

- How to improve the storage of geodata to keep customers satisfied in what they really need on information and advise in the future?
- How to facilitate data interconnexion on the web based on reference data?
- What is the perspective of optimal geodatastorage in the next years?
- What are architecture solutions to mediate INSPIRE and ELF compliant portals with Linked data world?
- Which investments does a NMCA take to get "linked-data-proof"
- Are there experiments a NMCA must start to get involved in knowledge about linked data?

In summary; linked data is a development which seems to connect to the field of research of EuroSDR members, gives partially answers to lots of actual questions of NMCA's and its users and is a topic which should be part of its activities. From this perspective, it's important for EuroSDR to keep involved in experiments and developments to improve its competences to keep its position as frontrunner in Geospatial data research. However, from a critical point of view, there are already al lot of standards for linked data standards and portals. Field of research by EuroSDR will be the position of a land registration and to develop an operational plan with needed resources to design needed linked data.

The conclusion is that there are a lot of relations to use linked data as a tool to connect content and data of different users and sources. From this perspective, there is a strong link with Commission 1 (sensor information en geo referencing). In the Netherlands, this aspect had an own track in PiLOD. Next to this aspect, there's a full relationship of linked data with Commission 5 (Methods and mechanisms for integrating core (framework) data with other geoinformation and business (or value-added) data, both by data linking and by interoperable data access). However, because of the strong interaction with Commission 4 from this perspective, the full subject linked data could be related to commission 4 as well.

5. Conclusions and recommendations

Conclusions

The question was what should be the actual role EuroSDR should play in the future in relationship to EuroSDR's activities and stakeholders. Next conclusions will answer this question:

Linked data is a development which seems to connect to the field of research of EuroSDR members in two ways :

- First, it's another way to address similar issues (distributing and sharing information on the web) It is worthwhile studying if these technologies perform better than the state of the art techniques used by NMA for some issues like multiple representation, schema transformation, cataloguing;
- Second, it yields new needs for location framework –different contents from INSPIRE contents- and it is important to have a uniform location framework across INSPIRE and the web of data. In this context, EuroSDR strength relates to: our competence in data integration

based on spatial criteria, our competence in heterogeneities encountered while modelling reality Europe-wide.

Linked data as research topic should be part of Commissions 4 or 5;

From the aspect of georeferencing, this element of linked data is part of Commission 1 research as well;

Linked Open Data may also give us interesting clues about the funding of information infrastructures.

Linked Data gives partially answers to lots of actual questions of NMCA's in data storage in future and its users and is a topic which should be part of its activities.

EuroSDR must beiing involved in experiments and developments and has to improve its competences to keep its position as a frontrunner in Geospatial data research.

Recommendations
Some recommendations could be defined:

EuroSDR could play a role in bringing experiments and activities together. The topic Linked Data should be adopted in one of the sub-commissions, proposaled in number 4 or 5;

Field of research by EuroSDR will be the position of a land registration and to develop an operational plan with needed resources to design needed linked data;

During next EuroSDR delegates meetings, linked data will be topic at the meeting schedule. Common activities in linked data will be reported an initiated;

During the 125th delegate-meeting in Leuven in Oktober 2014 it was decided EuroSDR will organise a Linked Data expert meeting with non- and members in spring 2015. Some issues can be shared, presented an solved. After that, a course can be initiated to get people more involved in the theory about linking data.

Reference list

- Folmer, E., et all.; "Pilot Linked Open Data, Deel 1", Geonovum Amersfoort 2013.
- Foler, E., et all.; "Pilot Linked Open Data, deel 2, De verdieping", Geonovum Amersfoort 2013.

Website

Website Geonovum: www.geonovum.nl/onderwerpen/linked-data/nieuws/