# Opportunities and Challenges for using Linked Data in INSPIRE

Sven Schade, Michael Lutz

European Commission – Joined Research Centre
{sven.schade, michael.lutz}@jrc.ec.europa.eu

**Abstract.** In this statement of interest, we investigate opportunities to use a Linked Data approach in INSPIRE (Infrastructure for Spatial Information in Europe). We briefly present two typical use cases related to INSPIRE and identity the involved components, as well as their workflow. In a subsequent step, we analyze the potential of using Linked Data for each use case. We conclude with a discussion on possible future analysis and implementation work. These specifically include the development of link types relevant for the presented INSPIRE use cases.

**Keywords:** Linked Data, Spatial Data Infrastructure, INSPIRE

## 1 Introduction

*Linked Data* has recently received a lot of attention as a means for making the tacit connections between online-resources explicit and for advanced information browsing [1],[2]. It refers to a best practice for exposing, sharing, and connecting resources, such as textual documents, pictures and maps, in the (Semantic) Web [1]. It is based on (i) the use of Uniform Resource Identifiers (URIs) [3] as reference points, (ii) the Resource Description Framework (RDF) [4] as basic structure for any form of description, and (iii) content-negotiation [5] to allow a client to specify an acceptable representation.

While numerous systems and applications for the environmental sciences have recently been based on the Linked Data approach, most of these were missing concrete use cases [6]. In this paper, we investigate the opportunities to use a Linked Data approach in the context of Spatial Data Infrastructure (SDI), which provides one concept supporting interoperability within and across involved information communities [7]. From an SDI development perspective, using Linked Data for integration purposes seems a logical next step [8]. We specifically focus on INSPIRE (Infrastructure for Spatial Information in Europe), which is one of the few SDIs with a legal basis [9]. We analyze example use cases for INSPIRE, identify research questions that need to be addressed in order to apply Linked Data in INSPIRE, and indicate future steps.

The following INSPIRE components are relevant for the work at hand:

- INSPIRE data specifications[1] which specify harmonized data models for 34 spatial data themes.
- INSPIRE Feature Concept Dictionary (FCD)[2] which is part of the INSPIRE registry[3] and includes natural language definitions of the concepts underlying

[1] http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2/
[2] https://inspire-registry.jrc.ec.europa.eu/registers/FCD/

the feature types defined in the INSPIRE data specifications. Similar registers are planned for other artifacts, such as code lists.

- INSPIRE-compliant metadata [10] on data sets which can be in any of the 23 official languages of the EU.
- GEneral Multilingual Environmental Thesaurus (GEMET)[4] which is mandated in the metadata Regulation as a controlled vocabulary for a keyword describing the spatial data theme.
- INSPIRE network services [11], which enable the access to data and metadata in the infrastructure.

The remainder of the document is structured as follows. Possible use cases are outlined in the next section. They illustrate possible requirements for the use of Linked Data in the context of INSPIRE. Especially, different links types are requested. Related issues will be discussed in the last section of this paper. Here, we also include research questions, and provide pointers to related work.

## 2 Two INSPIRE Use Cases

We selected two representative use cases from INSPIRE; data discovery (described in more detail in [12]) and cross-theme queries. Both will be briefly described in the following, including the involved INSPIRE components and basic workflows.

**Discovery.** A user wants to find a specific data set using a client to an INSPIRE discovery service. He starts entering keyword (in a search language), for example ‘coast’, which is auto-completed by the client based on a multi-lingual controlled vocabulary (mlCV), such as GEMET. The client expands the query based on the relationships in the mlCV in order to find additional search terms (‘cliff’, ‘cove’, and ‘seashore’), and (optionally) some of these terms are selected by the user. Based on selected terms, a catalogue query is generated, in which the terms’ translations in a number of target languages are compared with the keyword elements in the metadata.

Alternatively to GEMET, the FCD can be used. In this scenario, the user enters a concept from the FCD (for example ‘shore’), and the query expansion is based on the relationships within the FCD, or between concepts in the FCD and concepts in other thesauri (GEMET etc.). In the former case, in order to find a match, the metadata record for a given data set would need to include the information which INSPIRE feature types are contained in the data set.

Once data sets that match the query have been identified a corresponding download or view network service has to be discovered. If this service is directly linked (as a *coupled resource*) from the metadata of the data set, the user can ‘browse’ directly to it (in contrast to the ‘query’ steps illustrated above). Yet the user needs to know the name of the desired feature type or layer in order to be able to submit a valid service request.

[3] https://inspire-registry.jrc.ec.europa.eu/

[4] http://www.eionet.europa.eu/gemet/

**Cross-theme queries.** The INSPIRE data models contain a number of cross-theme relationships, for example that between an address (represented as points) and the transport link (for example a road) that forms part of the address[5]. Thus, there can be relationships between objects in two (or more) spatial data sets, possibly being made available using different INSPIRE download services. Like in this example, these relationships cannot in all cases be derived from spatial characteristics (an address near a cross road might not contain the road it is closest to).

In the use case, a user wants to query INSPIRE data sets based on the links that exist between them. For example, he might be interested in finding all addresses that are associated with a specific road. For such a query, the relationships between the spatial objects must be available and query-able, potentially across several INSPIRE download services.

## 3 Discussion and Preliminary Conclusion

When reconsidering the three main principles of Linked Data, these are already partly provided for in INSPIRE. Earlier investigations illustrated that linking could be equally achieved using common geospatial/SDI standards [13]; especially as the Open Geospatial Consortium (OGC)[6] just supports the http URI scheme and the OGC naming authority provides their dictionaries in http, GML [14], and RDF encoding. In INSPIRE, URIs as reference points could be derived from the *inspireId* attributes of spatial objects or from the relevant INSPIRE registers (for example for feature types, feature concepts or data sets). Currently, GML is specified as the default encoding for all existing INSPIRE data specifications. However, since GML and RDF are isomorphic [13], the data could be provided in RDF as well. The decision of using a GML or RDF representation of geospatial data depends on the intended use. Both are optimized for different purposes. While RDF allows for sophisticated querying and reasoning, numerous GIS clients process GML. In general, equipping geospatial web services with content negotiation would be beneficial in general.

This suggests that implementing INSPIRE requirements using a Linked Data approach is feasible. However, a number of open questions still remain.

**Added benefit.** From a general perspective, linking INSPIRE resources to existing (third-party) data clouds, such as Linked Geodata[7], is desirable. But, on which level could we benefit from a Linked Data implementation? Certainly Linked Data would have the benefit of enabling the use of existing Linked Data tools and the development of rich clients. It may also be useful in reporting (SEIS[8] provides an example), where the semantics of links are well known, but is this true for SDI in general? Do we want to introduce MIME types for each INSPIRE theme? Where would INSPIRE

---

5 See https://inspire-twg.jrc.ec.europa.eu/inspire-model/index.html?goto=2:3:1:7266 for the cross-theme relationships of the theme *Addresses*.

6 http://www.opengeospatial.org/

7 http://linkedgeodata.org/

8 http://ec.europa.eu/environment/seis/

benefit from ontologies on top of plain RDF, which may offer easy navigation and inference of new knowledge?

**Browsing vs. Querying.** As illustrated in the use cases, there are two prominent interaction pattern with resources in INSPIRE and ways to make use of the links between them: The 'browse' pattern, where a user subsequently discovers additional facts by following links, and the 'query' pattern, where the user specifies clearly (using the query) what he is interested in. While the 'browse' pattern is directly supported by Linked Data, the 'query' pattern is at the basis of existing OGC service specifications. It needs to be investigated how the benefits of both approaches can be combined to achieve maximum benefit, especially for cross-resource queries.

**Links.** In order to implement the use cases, a number of specific link types would be required. For the query expansion in the discovery scenario, the SKOS[9] properties for relating concepts to concepts (in the same or different SKOS) and for relating concepts to labels can be used. But, is the SKOS vocabulary expressive enough? Which relations would be useful in addition? Other – INSPIRE-specific – link types (for example between metadata records and keywords, or between spatial objects belonging to specific feature types) could be derived from the existing legal texts and corresponding guidelines. However, there are other links which have not yet been defined in any of the INSPIRE documents, for example the link between a metadata record on a data set and the feature type(s) or layer(s) the data set provides. Should we develop INSPIRE ontologies (in the Semantic Web sense)?

How do we maintain links, especially in the metadata records and catalogues? How can link 'vocabularies' be shared across communities? How do we provide mechanisms to translate between such vocabularies? Especially this last point may benefit from research activities, such as 'ontology localization' [15], but how exactly?

**Metadata.** With Linked Data, the distinction between data and metadata becomes less pronounced. In principle, every linked information item could be considered as metadata on the data which links to it. Thus, when applying Linked Data principles to INSPIRE, the role of metadata (records) and discovery services as the central repository for metadata would need to be investigated. This includes elaborations on the use of linked metadata for data sets, data objects, layers, and services, as well as inclusion of metadata links into delivered data sets and maps.

**URIs and Registers.** Governance and management of identifiers is an important issue to be addressed. When National SDIs join the Linked Data Initiative and they generate their own URIs, will we face issues duplication? Interesting example for a strategy on URI management can be found in [16] and [17].

As discussed above, INSPIRE registers could provide some of the required URIs, for example for feature types or code list values. Currently the INSPIRE registry is geared towards human consumption. In order to enable their usage in a Linked Data architecture, they should also offer information in a machine-process-able form.

[9] http://www.w3.org/TR/skos-reference/

In summary, Linked Data seems to connect the Semantic Web and geospatial, and especially SDI community closer to each other. Benefits for users seem obvious; they can build mash-ups and applications on top of INSPIRE data more easily. For this purpose, proper links between potential INSPIRE ontologies and data with existing geospatial data sources may be considered It still has to be clarified if this would be beneficial for the (legally binding) INSPIRE data sets and services. We are currently investigating most of these questions in research projects (GENESIS[10], EuroGEOSS[11], and GEO AIP3[12]). In addition we are active in the Ecoterm initiative, which aims at advancing provision and use of environmental terminologies. Complementary to these 'vertical' activities, we aim to broaden our investigations in Linked Data use cases in INPSIRE to ensure 'horizontal' coverage of opportunities and challenges.

## References


[1] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data: Principles and State of the Art," 2008.

[2] C. Bizer, "The Emerging Web of Linked Data," *IEEE Intelligent Systems*, vol. 24, 2009, pp. 87-92.

[3] W3C/IETF URI Planning Interest Group, "URIs, URLs, and URNs: Clarifications and Recommendations 1.0."

[4] O. Lassila and R.R. Swick, *Resource Description Framework (RDF) Model and Syntax Specification*, 1999.

[5] K. Holtman and A. Mutz, "RFC 2295 - Transparent Content Negotiation in HTTP", Mar. 1998.

[6] G. Hodge, *Report on the Outcome of the ECOTERM V Workshop*, U.N. Food and Agriculture Organization, Rome, Italy: 2009.

[7] H. Onsrud, *Research and theory in advancing spatial data infrastructure concepts*, Redlands Calif.: ESRI Press, 2007.

[8] S. Schade, C. Granell, and L, Díaz, "Augmenting SDI with Linked Data", *GIScience 2010 Workshop - Linked Spatiotemporal Data*, 2010, accepted for publication.

[9] INSPIRE, *INSPIRE Scoping Paper*, 2004.

[10] INSPIRE, *INSPIRE Metadata Regulation*, 2008.

[11] INSPIRE, *Network Services Architecture (Version 3.0).*, 2008.

[12] M. Lutz, N. Ostländer, X. Kechagioglou, and H. Cao, "Challenges for Metadata Creation and Discovery in a multilingual SDI – Facing INSPIRE," Stresa, Italy: 2009.

[13] S. Schade and S. Cox, "Linked Data in SDI or How GML Is Not about Trees," Guimarães, Portugal: 2010.

[14] OGC, *Geography Markup Language (GML) Encoding Standard v3.2.1*, Open Geospatial Consortium Inc., 2007.

[15] M. Espinoza, E. Montiel-Ponsoda, and A. Gómez-Pérez, "Ontology Localization", 5th *Fifth International Conference on Knowledge Capture (KCAP)*, 2009.

[16] Paul Davidson, *Designing URI Sets for the UK Public Sector*, UK Chief Technology Officer Council, 2009.

[17] D. Ayers and M. Vlkel, "Cool URIs for the Semantic Web", *Interest Group Note 20080331*, W3C, 2008.


---

[10] http://genesis-fp7.eu/

[11] http://www.eurogeoss.eu/

[12] http://www.ogcnetwork.net/node/628/