

## **Cross-Domain Metadata Interoperability: Lessons Learnt in INSPIRE**

Andrea Perego

Michael Lutz

Massimo Craglia

Silvia Dalla Costa

European Commission – Joint Research Centre, Italy  
{firstname.lastname}@jrc.ec.europa.eu

### **Abstract**

Since 2007, EU Member States have been involved in creating an infrastructure for spatial information in Europe (INSPIRE), based on a legal and technical interoperability framework. This paper presents some of the lessons learnt during the implementation of this infrastructure (which started in 2009) and during work on data and service interoperability coordinated with European and international initiatives. We describe a number of critical interoperability issues affecting both scientific and government data and metadata, and propose how these problems could be effectively addressed by a closer collaboration of the government and scientific communities, by taking advantage of their complementary competencies, and by influencing the development and adoption of standards.

**Keywords:** INSPIRE; metadata; interoperability; multilingualism; licensing; data quality; provenance; lineage; versioning; persistent identifiers; harmonisation; standardisation.

### **1. Background**

INSPIRE is a Directive (OJ, 2007)<sup>1</sup> of the European Parliament and of the Council aiming to establish a EU-wide spatial data infrastructure to give cross-border access to information that can be used to support EU environmental policies, as well as other policies and activities having an impact on the environment. The actual scope of this information corresponds to 34 environmental themes<sup>2</sup>, covering also areas of cross-sector relevance – e.g., addresses, buildings, population distribution and demography.

In order to ensure cross-border interoperability of data infrastructures operated by EU Member States, INSPIRE sets out a framework based on common specifications for metadata, data, network services, data and service sharing, monitoring and reporting. Such specifications consist of a set of implementing rules (i.e., legally binding legislation), along with the corresponding technical guidelines.

The datasets, dataset series and services that make up the INSPIRE infrastructure can be discovered based on harmonised metadata and catalogue services (called “discovery service” in INSPIRE) giving access to this metadata. The INSPIRE metadata schema is defined in the INSPIRE Metadata Regulation (OJ, 2008), and it includes a number of elements relevant for resource discovery. Further metadata elements for evaluation and use are defined in the INSPIRE Regulation on interoperability of spatial data sets and services (OJ, 2010). Some of the key features include support to cross-language and spatial search and semantic annotations based on controlled vocabularies and thesauri.

Following the INSPIRE implementation roadmap<sup>3</sup>, since December 2010 EU Member States are making available INSPIRE metadata, which since November 2011 have to be published through INSPIRE discovery services. These discovery services are used by the INSPIRE

---

<sup>1</sup> The full list of the legal and technical documentation concerning INSPIRE is available from the INSPIRE Web site: <http://inspire.ec.europa.eu/>

<sup>2</sup> For the list of INSPIRE themes, see: <http://inspire.ec.europa.eu/index.cfm/pageid/2/list/7>

<sup>3</sup> See: <http://inspire.ec.europa.eu/index.cfm/pageid/44>

Geoportal<sup>4</sup>, operated by the European Commission, to harvest and index metadata, thus providing a single access point to discovery INSPIRE data and services from EU Member States.

It is worth noting that, although INSPIRE focuses on environmental data, the majority of INSPIRE metadata elements are generic enough to describe also other types of data and services (a summary of the elements of the INSPIRE metadata schema is provided in Figure 1). Moreover, INSPIRE metadata are currently being used for describing data and services of both public administrations and research organisations.

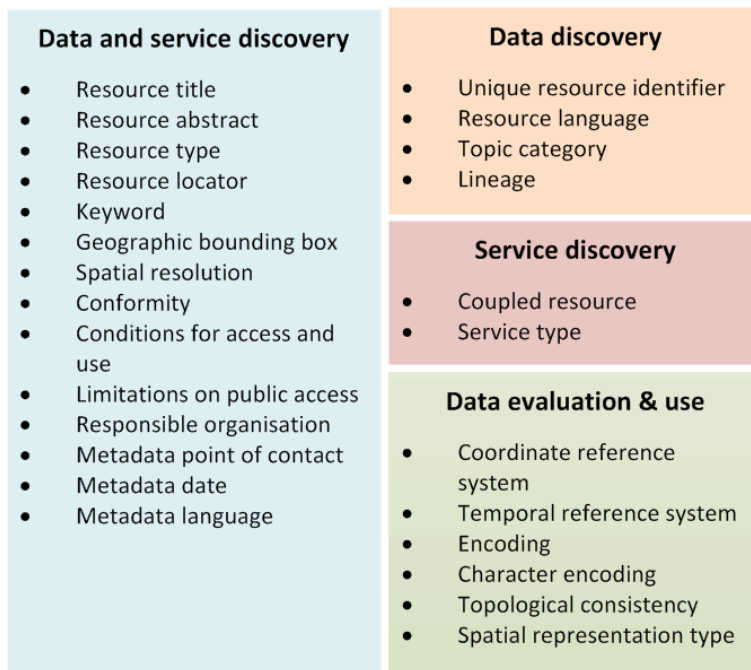


FIG. 1. INSPIRE metadata elements at a glance

For these reasons, the Joint Research Centre of the European Commission (JRC), as technical coordinator of the implementation of INSPIRE, is collaborating in a number of European and international initiatives concerning best practices and cross-domain interoperability of government and scientific metadata (for a general overview, see Perego et al., 2012). These initiatives include a number of working groups chartered in the framework of Research Data Alliance<sup>5</sup>, and work in the European Commission on the definition of core vocabularies for public administrations<sup>6</sup>, a vocabulary and framework for sharing semantic interoperability assets<sup>7</sup> and of a common metadata interchange format for European data portals, based on the Data Catalog Vocabulary of W3C<sup>8</sup>. At the same time, JRC is investigating possible extensions to INSPIRE metadata in order to address domain-specific requirements for scientific data, in the scope of a number of activities concerning EU institutions and Member States.

## 2. Outstanding interoperability issues

One of the main lessons learnt in INSPIRE and in the initiatives JRC is involved, is that government and research data are not two separate worlds. Although they may have different and

<sup>4</sup> The INSPIRE Geoportal is available at: <http://inspire-geoportal.ec.europa.eu/>

<sup>5</sup> See: <https://rd-alliance.org/>

<sup>6</sup> See: [http://joinup.ec.europa.eu/community/core\\_vocabularies/description](http://joinup.ec.europa.eu/community/core_vocabularies/description)

<sup>7</sup> See: <http://joinup.ec.europa.eu/asset/adms/description>

<sup>8</sup> See: [http://joinup.ec.europa.eu/asset/dcat\\_application\\_profile/description](http://joinup.ec.europa.eu/asset/dcat_application_profile/description)

domain-specific requirements, their scopes are often overlapping. Furthermore, government data are commonly used as a basis to create scientific data, and vice-versa. Consequently, it is fundamental to adopt a consistent approach to address interoperability issues shared by both government and scientific data.

In particular:

- Controlled vocabularies are a key component to support semantic and cross-language data discovery across domains. However, what is missing is a framework for the (collaborative) maintenance and publication of controlled vocabularies, providing support to multilingualism (cross-language indexing and search), versioning (backward / forward interoperability) and mapping - both between terms in the same vocabulary and in different ones (cross-domain interoperability).
- The lack of a consistent approach to licensing is one of the key issues that prevents effective data re-use. Possible solutions include transparent and machine-readable representations of licences for data and services, and a common protocol to digital rights management, for resource discovery, access, and use.
- The ability to verify the quality of data is fundamental both in a government and scientific context. This may be addressed by adopting transparent and machine readable representations of data provenance, lineage, and use. Other features may include support to users' feedback and quality rating (possibly also through third-party quality certification).
- Data may be subject to changes and updates, therefore support to data versioning would grant access to historical data. This is important, for instance, when decisions or predictions based on given data result to be incorrect. In such cases, the ability to restore the original data is fundamental. The widespread use of persistent identifiers may play an important role also to address this issue.

Notably, the government and research communities have specific competencies that can be re-used to address some of such issues. An example is the long and consolidated tradition concerning data archiving and curation typical of the scientific community, which can be adapted to government data, without re-inventing the wheel.

Moreover, such issues are currently addressed by a number of initiatives, most of them running in parallel, sometimes ignoring relevant work carried out in other communities and often proposing not interoperable or conflicting solutions. The government and scientific community could create a critical mass that might promote a better coordination towards the definition and adoption of effective and consistent solutions for cross-domain interoperability.

## **References**

- OJ. (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). OJ L 108, 25.4.2007, p. 1–14.
- OJ. (2008). Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata (Text with EEA relevance). OJ L 326, 4.12.2008, p. 12–30.
- OJ. (2010). Commission Regulation (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services. OJ L 323, 8.12.2010, p.11.
- Perego, Andrea, Cristiano Fugazza, Lorenzino Vaccari, Michael Lutz, Paul Smits, Ioannis Kanellopoulos, and Sven Schade. (2012). Harmonization and Interoperability of EU Environmental Information and Services. *IEEE Intelligent Systems*, 27(3):33-39, May-June 2012. doi: 10.1109/MIS.2012.22