RDF and PIDs for INSPIRE: a missing item in ARE3NA

Robin S. Smith[1], Michael Lutz[1], Andrea Perego[1], Anders Friis-Christensen[1], Alice Vasilescu[2], João Rodrigues Frade[2], Danny Vandenbroucke[3] and Diederik Tirry[3]

[1] European Commission – Joint Research Centre (JRC)
[2] Deloitte
[3] University of Leuven (KU Leuven)

Introduction

Our proposed presentation will outline intermediate results of a study in the context of geospatial data sharing across borders and at European level. The study is aiming to develop a common approach to generating common RDF schemas for representing INSPIRE data and metadata, as well as guidelines for the governance of persistent identifiers (PIDs). These are important elements for enabling the re-use of INSPIRE data in other sectors, in particular in e-government. The results of the study may feed into a proposal for additional encoding rules and guidelines for INSPIRE and will be performed in close collaboration with the INSPIRE Maintenance and Implementation Group and the ISA Programme's Spatial Information and Services Working Group.

INSPIRE and the EU ISA Programme

By providing a legal and technical framework for a EU-wide data infrastructure, INSPIRE is the first effort to share geospatial data that has been undertaken across Europe. As a directive agreed upon by the EU Member States, it legally requires public administrations to provide access to the data sets in the Directive's scope, in part indicated by the 34 themes contained in its three annexes (see Figure 1).

Annex I	**Annex III**
1. Coordinate reference systems	14. Statistical units
2. Geographical grid systems	15. Buildings
3. Geographical names	16. Soil
4. Administrative units	17. Land use
5. Addresses	18. Human health and safety
6. Cadastral parcels	19. Utility and governmental services
7. Transport networks	20. Environmental monitoring facilities
8. Hydrography	21. Production and industrial facilities
9. Protected sites	22. Agricultural and aquaculture facilities
	23. Population distribution – demography
	24. Area management/restriction/regulation zones & reporting units
Annex II	25. Natural risk zones
	26. Atmospheric conditions
10. Elevation	27. Meteorological geographical features
11. Land cover	28. Oceanographic geographical features
12. Ortho-imagery	29. Sea regions
13. Geology	30. Bio-geographical regions
	31. Habitats and biotopes
	32. Species distribution
	33. Energy Resources
	34. Mineral resources

Figure 1: Data themes of INSPIRE

Although INSPIRE has a focus on supporting the implementation and evaluation of European environmental policies, reflected in many of the themes, it takes a wide view on the data necessary, ranging from reference data themes (e.g. addresses or cadastral parcels) to environmental themes (e.g. soil or species distribution). It also covers data themes with objects that may have an impact on the environment (e.g. production and industrial facilities) and that may be impacted (e.g. human health and safety or population distribution).

INSPIRE lays down legal obligations in so-called Implementing Rules (IRs), which are directly applicable in the EU Member States. They are complemented with Technical Guidelines (TGs) containing detailed instructions and recommendations for implementers. While IRs specify *what* needs to be implemented at an abstract and generic level, the non-binding TGs specify *how* to implement it, making reference to existing standards where appropriate.

A study on RDF and PIDs for INSPIRE

One of the main areas of INSPIRE is addressing the interoperability of geospatial data sets and services through harmonised data models and encodings for the exchange of data related to one of the 34 spatial data themes listed in the Annexes to the INSPIRE Directive. These data models have been developed on a conceptual level using the Unified Modeling Language (UML). The default encoding recommended to be used in the INSPIRE Technical Guidelines is automatically generated from these UML data models based on explicitly defined encoding rules. The current default encoding for most INSPIRE themes is based on the Geography Markup Language (GML)[1]. While GML is widely known within the geographical information domain, other e-government applications and tools are made available as Linked Data using Semantic Web technologies, as Resource Description Framework (RDF)[2]. A number of on-going initiatives in EU Member States (e.g. [4-7]) and EU projects (such as InGeoCloudS, GeoKnow [1-2]and SmartOpenData) are creating RDF vocabularies based on the INSPIRE data models. However, there are currently no agreed rules or guidelines on how to create such RDF vocabularies from INSPIRE data models, and consequently the derived RDF vocabularies (and hence the RDF data represented with them) are likely to be incompatible. This problem, while most prominent for the interoperability of spatial data, also applies to metadata, services and registries.

A related problem is the issue of global persistent identifiers (PIDs). Such identifiers would benefit the implementation of INSPIRE, in particular for items such as spatial data sets and services, individual spatial objects in these data sets, real world entities, code list values, documents, coordinate reference systems and others. As part of their eGovernment initiatives, several EU Member States (e.g. [3,8]) have also started creating governance structures, processes, rules/guidelines and tools to create, manage, maintain and use such PIDs in their Spatial Data Infrastructures (SDIs) that INSPIRE is built upon. Again, different approaches have been taken in different EU Member States and it would be beneficial to identify best practices and guidelines in this area to aid the governance of PIDs across the EU, for the geospatial/location community and others dealing with data-sharing in many sectors of the public administration, in private companies as well as in research institutions.

What we want to contribute to, and learn from, the workshop

In our presentation, we will give an overview of a study that is looking into these two issues, with the aim of creating

[1] See: http://www.opengeospatial.org/standards/gml
[2] See: http://www.w3.org/RDF/

- an agreed, common methodology for developing INSPIRE RDF vocabularies based on the UML data models defined for the INSPIRE Spatial Data Themes, and
- guidelines for governance models and processes for managing PIDs.

The presentation will focus on the current state-of-play in EU Member States, in specific projects and particular initiatives related to RDF for location and location-related PIDs. This includes examining current and planned actions, good practices, methodologies and tools. Special attention will be paid to identifying and supporting the community active in these areas across Europe.

The LGD workshop will offer an opportunity to show what we have learned so far, gather further evidence (from meeting presentations and relevant sources from participants), build interest in the work and identify interested experts to provide examples of good practice, candidate methodologies and examples of technologies that could support this work, as well as any potential pitfalls or obstacles.

Furthermore, we hope to build on the expected momentum from the meeting to promote two virtual workshops planned for RDF and PID within the scope of INSPIRE, where interim findings will be discussed.

Acknowledgements

The work outlined below is part of the work programme of the ISA-funded Action, *A Reusable INSPIRE Reference Platform* (ARE3NA – ISA Action 1.17)[3], which aims to gather and share examples of technologies aiding INSPIRE implementation (especially readily reusable open source tools); identify missing items within INSPIRE and the 'interoperability bridges' between INSPIRE and other cross-border/cross-sector activities.

References

[1] GeoKnow project – Deliverable 2.1.1. Available at: http://svn.aksw.org/projects/GeoKnow/Public/D2.1.1_Market_and_Research_Overview.pdf

[2] GeoKnow project – Task 2.7: Exposing INSPIRE data as Linked Data. Task description available at: http://geoknow.eu/t2-7.html

[3] Hans Overbeek, Thijs Brentjens. Draft URI Strategy for the NL Public Sector In: Open Data on the Web, 23-24 April 2013, London. Available at: http://www.w3.org/2013/04/odw/odw13_submission_14.pdf

[4] Linda van den Brink, Paul Janssen, Wilko Quak: From Geo-Data to Linked Data: Automated Transformation from GML to RDF, http://www.pilod.nl/wiki/Boek/BrinkEtAl-GML2RDF

[5] Sonya Abbas, Adegboyega Ojo. Applying Design Patterns in URI Strategies - Naming in Linked Geospatial Data Infrastructure. In: Hawaii International Conference on System Sciences (HICSS-47), January 6-9, 2014 [to appear], IEEE Computer Society, Hilton Waikoloa, Big Island, United States, 2014.

[6] Sonya Abbas, Adegboyega Ojo. Towards a Linked Geospatial Data Infrastructure. EGOVIS/EDEM 2013: 196-210. DOI: 10.1007/978-3-642-40160-2_16

[7] Sven Tschirner, Ansgar Scherp, Steffen Staab. Semantic Access to INSPIRE - How to Publish and Query Advanced GML Data. In: Proceedings of the Terra Cognita Workshop on Foundations, Technologies and Applications of the Geospatial Web. In conjunction with the International Semantic Web Conference (ISWC2011), Bonn, Germany, October 23, 2011, pp. 75-87. Available at: http://ceur-ws.org/Vol-798/paper7.pdf

[8] UK CTOC. Designing URI Sets for Location. Version 1.0. May 2011. Available at: http://data.gov.uk/library/designing-uri-sets-for-location

[3] See: http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-17action_en.htm