EuroSDR

European Spatial Data Research

February 2019

„Preserving Semantics, Tractability and Evolution on a multi-scale Geographic Information Infrastructure: Cases for extending INSPIRE Data Specifications"

Report of EuroGeographics – EuroSDR workshop on INSPIRE Data Extension

Bénédicte Bucher, Dominique Laurent, Paul Janssen

Workshop Report

Published by EuroSDR

EUROPEAN SPATIAL DATA RESEARCH

COMMISSION CHAIRPERSONS:

Data Acquisition: Jon Mills, United Kingdom
Modelling and Processing: Norbert Haala, Germany
Updating and Integration: Jon Arne Trollvik, Norway
Information Usage: Bénédicte Bucher, France
Business Models and Operation: Joep Crompvoets, Belgium
Knowledge Transfer: Markéta Potůčková, Czech Republic

OFFICE OF PUBLICATIONS:

Bundesamt für Eich- und Vermessungswesen
Publications Officer: Michael Franzen
Schiffamtsgasse 1-3
1020 Wien
Austria
Tel.: + 43 1 21110 825200
Fax: + 43 1 21110 82995202

CONTACT DETAILS:

Web: www.eurosdr.net
President: president@eurosdr.net
Secretary-General: secretary@eurosdr.net
Secretariat: admin@eurosdr.net

EuroSDR Secretariat
Public Governance Institute
KU Leuven
Faculty of Social Sciences
Parkstraat 45 bus 3609
3000 Leuven
Belgium
Tel.: +32 16 37 98 10

The official publications of EuroSDR are peer-reviewed.

Table of Content

Bénédicte Bucher, Dominique Laurent, Paul Janssen

"Preserving Semantics, Tractability and Evolution on a multi-scale Geographic Information Infrastructure: Cases for extending INSPIRE Data Specifications"

Report of EuroGeographics – EuroSDR workshop on INSPIRE Data Extension

PRESERVING SEMANTICS, TRACTABILITY AND EVOLUTION ON A MULTI-SCALE GEOGRAPHIC INFORMATION INFRASTRUCTURE: CASES FOR EXTENDING INSPIRE DATA SPECIFICATIONS

Report of EuroGeographics – EuroSDR workshop on INSPIRE Data Extension

With 8 figures

Bénédicte Bucher[1, 2], Dominique Laurent[1], Paul Janssen[3]

[1] IGN-France

[2] LASTIG-COGIT, University Paris Est

[3] Geonovum

Index of Figures

1 INTRODUCTION

Grounding on data the planning and monitoring of European environmental activities is an important issue for the construction of Europe. Reusing geo-data already produced by members to monitor activities at the level of the country or region in order to perform monitoring and planning at the global level of Europe is a different but related issue. It is a strategy to achieve more consistency and trust between the different levels where environmental activities and policies are monitored. This strategy has led the European Commission to issue the INSPIRE directive–Infrastructure for Spatial Information in Europe- ; national public authorities should make their data reusable for Europe [European Parliament 2007][Video 1]. Reusing at the level of Europe national data is not straight forward. Even though we share a common European physical territory, especially in the domain of environment where many phenomena do not stop at borders, each country has its own specific way to design data about ground water, transport networks, population, land-use and air temperature. The European digital territory also has its boundaries.

The 'INSPIRE extensions' workshop was organized to foster exchanges between organizations in charge of implementing INSPIRE, with a focus on the topic of extending INSPIRE Data Specifications, to share their experiences, improve each one's knowledge and also the scalability of local initiatives. It was co-organised by EuroGeographics Knowledge Exchange Network on INSPIRE (aka KEN INSPIRE), EuroSDR Commission 4 (focused on information usage) and Geonovum, a public organization who is coordinating the Dutch National Spatial Data Infrastructure (NSDI) supported by the Dutch Cadaster and the Geological Survey of the Netherlands.

Complementary to the minutes dedicated to the KEN INSPIRE community, this report wishes to be more pedagogic to reach out an audience –scientists or developers- who may not be familiar enough with INSPIRE issues and problems but could find INSPIRE as a possible application domain for its scientific or technological field and, in the future, be willing to contribute to the progress of INSPIRE.

The two first sections summarise the motivation of INSPIRE and the reasons for designing INSPIRE Data Extensions. The section after presents approaches adopted by practitioners: inheritance on the one hand and adaptation on the other. The subsequent sections tackle issues related to INSPIRE implementation in general and not limited to Data extension.

All presentations of the workshop together with the minutes are available on the workshop website:
https://eurogeographics.org/calendar-event/workshop-on-inspire-extension-june-2017/

2 Context: INSPIRE Data Specification and motivations for extensions

Surveying a vast, dynamic and heterogeneous territory like Europe refers to quite a set of challenges in terms of data provision. We recall briefly initiatives that have addressed these challenges.

The first European initiative to make use of digital geographic information to ensure consistent monitoring of environmental resources was the program Coordination of Information on the Environment (CORINE) launched in 1985 that led to the creation of the European Environmental Agency (EEA) - https://www.eea.europa.eu/ - in charge of coordinating data production for European environment information [Masser and Crompvoets 2015]. Today, the Copernicus program targets the generation of products with full coverage of Europe based on state of the art remote sensing technologies and processing chains applicable on all territories in Europe.

Additionally, in September 2000, public authorities in charge of providing geodata at the scale of nations or regions -the national mapping agencies, cadastral and land registration authorities- gathered to form an independent international not-for-profit organization representing them at the level of Europe: EuroGeographics, https://eurogeographics.org/ . EuroGeographics has an operational focus and feeds the Geographic Information System of the European Commission with pan-European products for topography, administrative boundaries and height data.

More recently, the advances of digital technologies and the rise of national geodata portals led to defining a strategy to achieve a European spatial information infrastructure at a fine scale –finer than EuroGeographics products- reusing 'relevant' data throughout local, national and transnational levels in order to ensure consistency and trust between all levels, from the region to Europe [Masser and Crompvoets 2015]. Reducing acquisition costs was also an argument for the INSPIRE directive to require that data produced by national legal authorities be available and reusable for analysis done at the European level [European Parliament 2007]. By design, INSPIRE implementation implies a strong connection with the producers of legal data. In practice, these producers and other experts have worked together to design a multiscale geographic information infrastructure content model. Key challenges are:

- making the right compromise between expressiveness and tractability: expressiveness calls for preserving local specificities whereas tractability calls for using an homogeneous abstraction,
- scalability: addressing all relevant scales, in a consistent manner, including temporal scales.

The INSPIRE technical groups have produced non-binding technical guidelines describing common data models, code lists, map layers and metadata to be used when exchanging spatial data sets : INSPIRE Data Specifications[1]. Reasons given for extending INSPIRE Data Specifications are to address two different issues.

- Firstly, to enrich the information infrastructure with new themes relevant to European environmental activities that were not initially foreseen in the INSPIRE directive. In environmental studies, systemic approaches often are required, and data involved in official surveys and reports need to embrace new themes. Extending European information infrastructure to include new categories of information could facilitate for member states the design of the information infrastructure needed tomorrow to manage the European territories.
- Secondly, to enrich data with local nation-specific concepts at the level of a country to reach more consistency and connections between current INSPIRE products and other national data (see figure 1).

[1] https://inspire.ec.europa.eu/Technical-Guidelines/Data-Specifications/2892

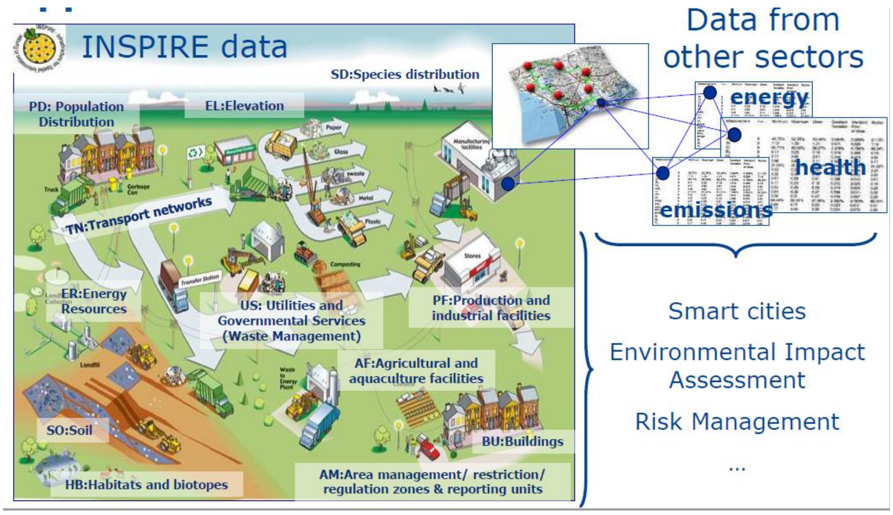

Figure 1: INSPIRE can be seen as a Base model in Europe to reference and interconnect to data from different sectors in the perspective of several applications like Smart cities, risk management © Lutz

Both issues, as underlined by Michael Lutz from the commission's Joint Research Center during the workshop, are comparable to those encountered by member states when working towards national-level information infrastructure, and favor data reusability across their country. As noted by Morten Borrebaek from Kartverket, national Norwegian mapping agency, using INSPIRE data specifications as a basis for national or community extensions is a way to enhance interoperability in digital information infrastructures, provided these extensions are designed in a consistent manner.

However, a questionnaire initiated by Geonovum and sent to INSPIRE user communities on 2016 revealed a lack of cooperation and coordination between communities extending INSPIRE, https://inspire.ec.europa.eu/news/survey-extending-inspire-data-models, [Grothe et al. 2016]. This led to the creation of the Inspire Data Specification Extensions project conducted by WeTransform : a platform to document and share best practices for extending INSPIRE models, http://inspire-extensions.wetransform.to/, and the design of a generic extensions design methodology that should ensure interoperability between models and data.

3 Comparing Technical Approaches to Implementing Data Extensions

3.1 Inheritance and INSPIRE extension rules

The standard mechanism to extend INSPIRE data specifications is to create subtype of existing data specifications following INSPIRE extension rules, cf. figure 2. The benefit is to ensure the conformity to INSPIRE.

This mechanism is for instance used in the official model that state member should use to provide data related to "Nationally designated areas" that are a specific category of protected sites in Europe (CDDA, 2018).

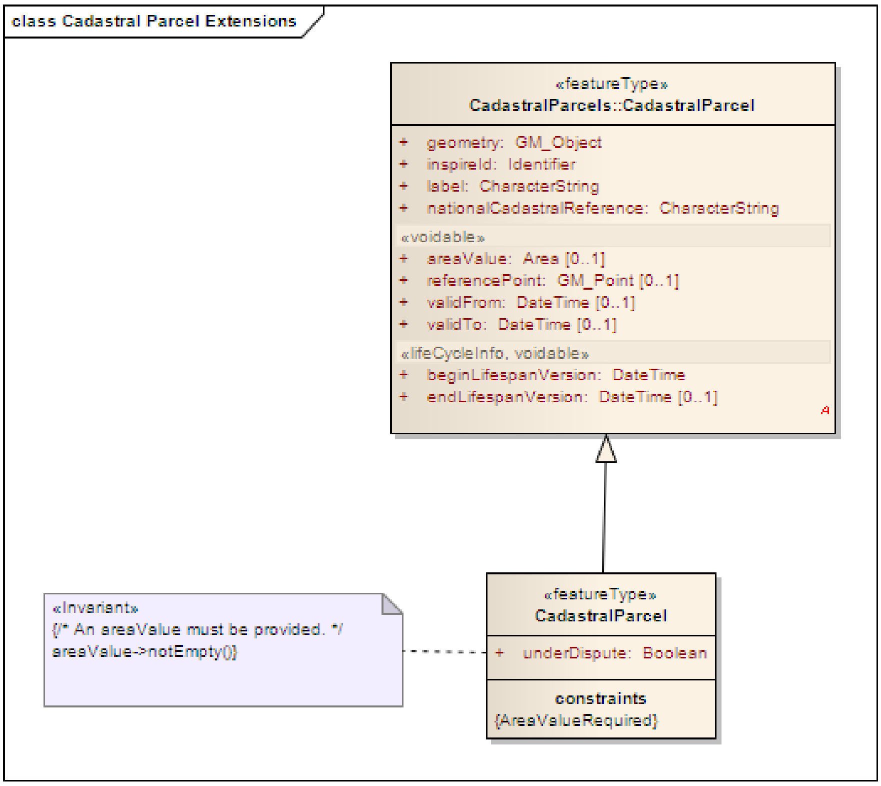

Figure 2: Use of inheritance mechanism (arrow between both featureTypes) to extend INSPIRE Data specifications and add information as a new attribute of an existing featureType [INSPIRE et al. 2013]

This approach has also been adopted by Eurogeographics in the European Location Framework (ELF) project. This project was funded by the Competitiveness & Innovation framework Programme (CIP) Information and Communication Technologies Policy Support Programme (ICT PSP) Open Data and open access to scientific information (Project reference 325140). It aimed at delivering a pan-European cloud platform and web services enabling access to INSPIRE geographical reference data at different levels of details from a single access point for the user, including for cross-border and pan-European data [Jakobsson 2012]. The data are not centralized in one data center but a mediation service is provided that interprets a user query and redirects it to the required national data services or to pan-European products data service –for regional and global levels of details-. The design of this new European platform required that all participants implemented extensions to INSPIRE specifications in an interoperable way, with the same data schema. The ELF project ended on October 2016 but the work is pursuing at EuroGeographics to turn ELF results into an operational service, European Location Service.

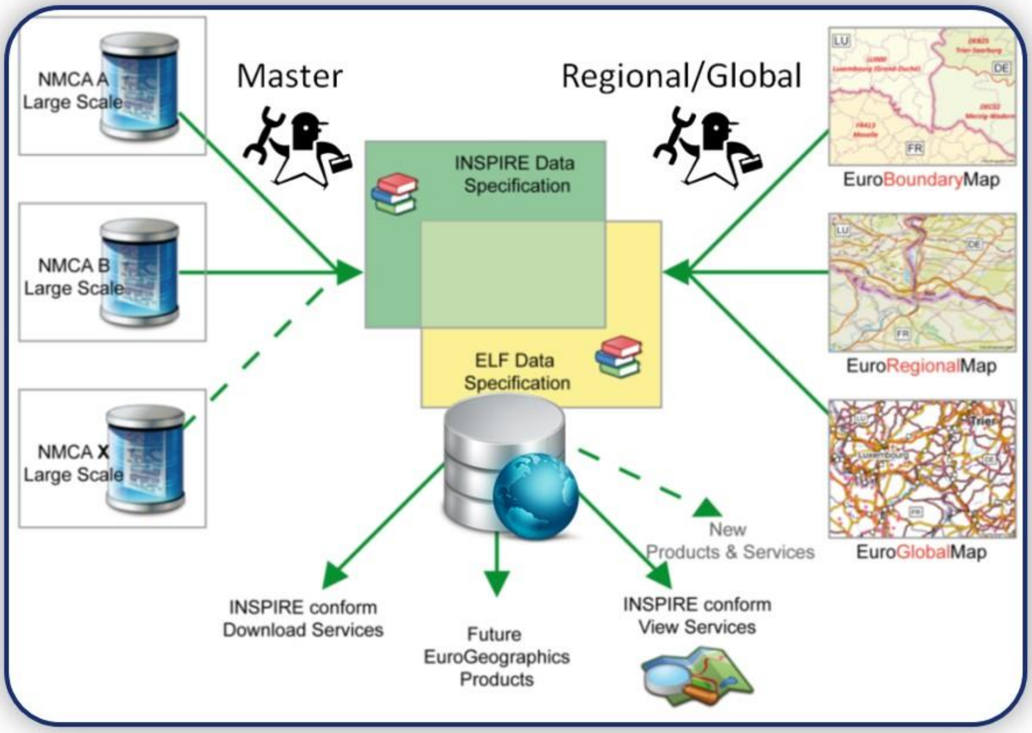

Figure 3: Extension of INSPIRE to connect to EuroGeographics pan-European products © Urbanas

The long term vision of EuroGeographics is to provide European Location Services that are single access point to authoritative data in Europe, connected with local data.

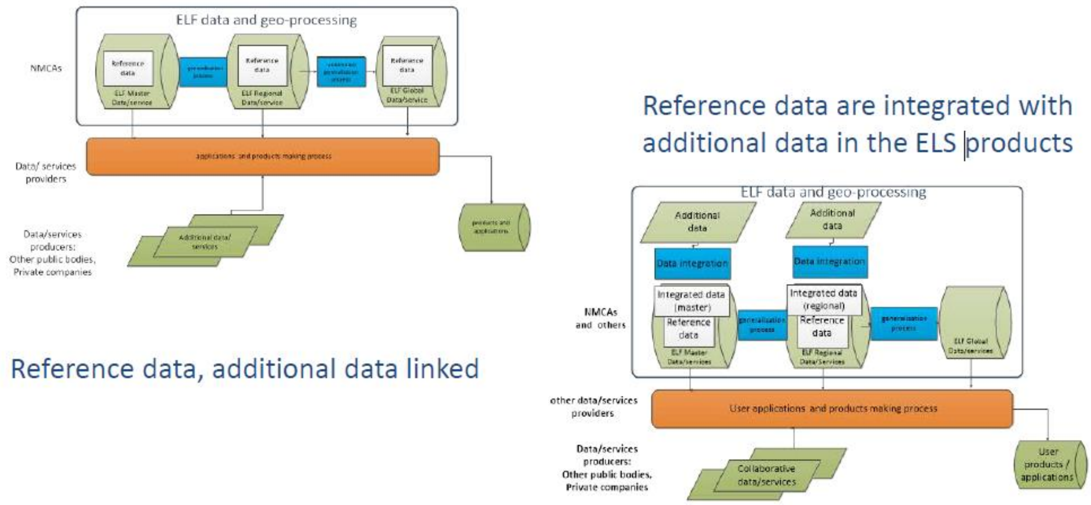

Figure 4: EuroGeographics long term vision for European Location Services © Urbanas

There are several issues with this extension mechanism given current database management technologies. New concepts necessarily inherit all attributes of the father concept, some of which are considered noisy/extraneous - for instance all voidable attributes if they are not filled in. Additionally, evolutions in the parent class impact all the inherited classes, which could be numerous. The usage of Linked Data technologies (publishing data via a series of triples consisting of object – predicate – subject to allow semantic queries) can overcome some of these issues, by organizing data in a distributed architecture. This approach was trialed by EEA for the next generation of Common Database for Designated Areas (CDDA) presented by Stefania Morrone from Epsilon, an Italian company providing solutions in the geospatial domainand in particular INSPIRE implementation. In this approach, the data corresponding to the INSPIRE model are provided through a INSPIRE download service (WFS) and the additional elements are stored as tabular data together with a URI referring to the INSPIRE data. All these data services are referenced in a EU Registry collecting identification and administrative data on facilities, installations and plants.

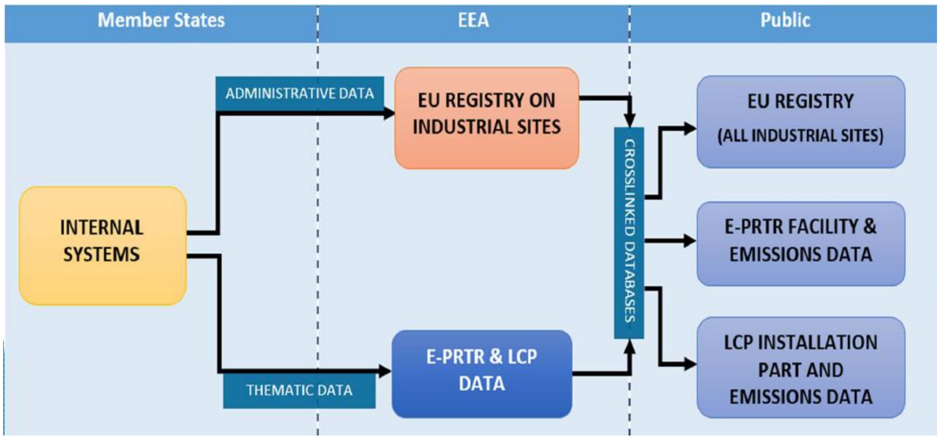

Figure 5: EURegistry on Industrial emissions © Morrone

4 Adaptations

An alternative mechanism to extend INSPIRE is referred to as an “adaptation”. Within this mechanism, two categories can be identified – redefinition/multiple inheritance and linking/realization.

A first category is similar to inheritance: redefine or multiple inheritance. An example is the alignment of the Dutch model for Cables and Pipelines Information, KLIC (Kabel en Leidingen Informatiecentrum), to INSPIRE presented by the Dutch Kadaster (figure 6). The reason to choose multiple inheritance is to keep the original KLIC model visible.

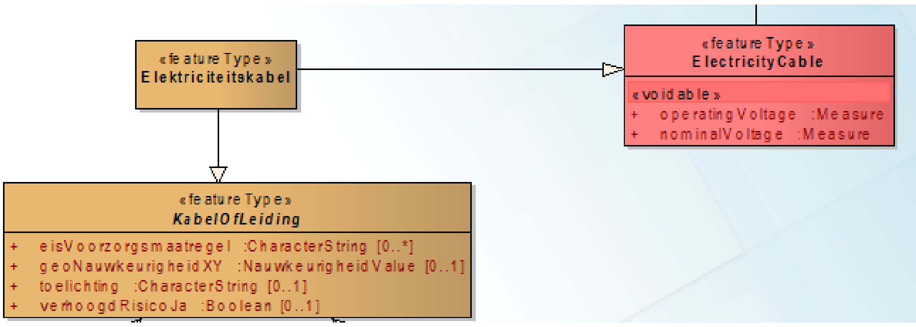

Figure 6: Aligning the Dutch cable and pipeline (KLIC) model to INSPIRE thanks to multiple inheritance, the new featureType Elektriciteitskabel inherits from INSPIRE ElectricityCable and from KLIC KabelOfLeiding © Van Houtum

A second category of methods for adaptation is to design a link between an existing concept and the related INSPIRE Data Specification, which can be compared to either with a realization pattern or with classical association.

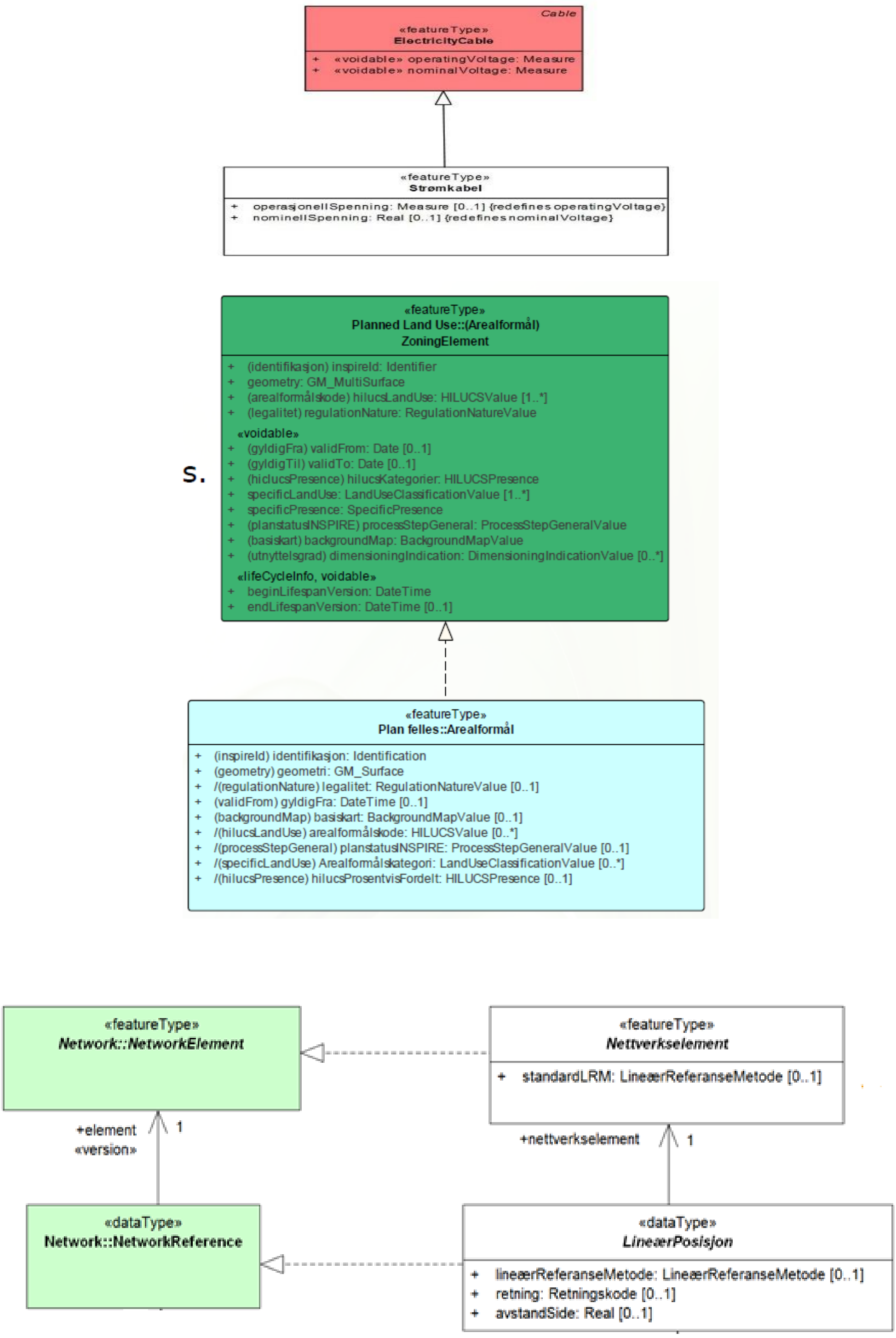

Figure 7: The Stromkabel featureType redefines INSPIRE ElectricityCable featureType (a). This pattern was not adopted by Karteverkert that preferred the realization pattern (b) and (c) © Borrebaek, Jetlund

An example of realizations was given by the Norwegian Mapping Authority –Kartverkert. The update of the national conceptual model for environmental information, SOSI (SOSI (Samordnet Opplegg for Stedfestet Informasjon), targeted among other objectives the alignment with INSPIRE. As shown in Figure 7 they also considered a redefinition method (7.a). The redefinition method from UML2.4 is a much valuable mechanism in theory (as it allows, for example, additional constraints to be added to the object) but proved to be too complex from a technical point of view: too little tool support. It was also considered dangerous because it can turn mandatory data to optional. In contrast, the realization method (objects in the target class support methods from the source class) does not have these weaknesses, and this approach was adopted for different themes like LandUse or Networks, figure 7 (b).

As a conclusion, adaptations are more suitable than inheritance to situations where production context requires different languages or requires flat models. Another key reason to choose an ad hoc adaptation solution is the necessity to reuse several INSPIRE application schemas together, e.g. PhysicalWaters and HydroNetworks. Linked data technologies are especially useful to support such linkage.

5 User adoption and reaching a sustainable economic model with industry

Besides opting for one or another technical approach to implement an INSPIRE Data Extension, a difficult case faced by most data providers is convincing stake holders to implement the extension at all and convincing industry to develop software solutions to valorize these data. Indeed, assessing and accessing the added value brought by an information infrastructure like INSPIRE is a challenge. There is to date no obvious solution to somehow monetize the added value of a new piece of information, for example local semantics, in a SDI. The capacity to combine efforts with the industry who may provide more commercial applications making use of the data is also hindered by the different strategies related to opening data. Besides, the challenge is all the more difficult than there are so many industries issuing standards -GIS, Web, Space, Building industry-. Several participants also reported that European applications and national implementations and legal requirements do not facilitate the INSPIRE vision of reusing at the European level data required for the national level; there may not exist a simple derivation mechanism to infer data required at European level from data required at national level –or vice versa-. Longer term, there is some uncertainty even among participants related to EC design choices to whether the different European directorates will actually use such structured semantically rich data which are, however, still very heterogeneous or whether they would prefer another technical solution based on rawer data such as that provided by Copernicus program even if –or because- they lack local semantics.

INSPIRE approach requires sufficient skills to identify and use data and their documentation. Data are at the same time homogeneous because they refer to a same conceptual model but very heterogeneous because they made different implementation choices. In some cases, users simply don’t know INSPIRE principles. In others they don’t know how to collect information about all national access points and specific APIs to access the data. This difficulty has led to the European Location Service approach.

To improve user adoption, workshop participants have developed several approaches to get closer to users and needs: looking at the law -what observations are required by law-, modeling the planning decisions process, designing data models oriented towards indicators such as the pilot project creating an EU disaster damage and loss database, studying application domain information content and requirements for spatial references –e.g. biodiversity or cables and pipelines - and requirements for spatial referencing frameworks. Geonovum also proposed to design an attention-grabbing use case (e.g. in the domain of energy) to explain and make more visible the added value of INSPIRE. Another newly formed initiative from the EuroSDR community to improve user-oriented cataloguing of NMCAs data is a 'geodata-advisor' to support exchange between people from application domain –e.g. climate change scientists- who can better express their expected benefit from spatial information and people who are experts in the technologies that describe available data and libraries –e.g. specialists of INSPIRE data, specialists of Sentinel data, specialist of photogrammetric libraries and so forth.

6 Getting enough expertise/assistance in digital information technologies, making design choices with enough awareness of (all) their impacts

Many presentations highlighted issues of the need within a NMCA for specialist expertise across the data workflow; it is not possible for one person to be expert in all required domains encountered during production. Numerous design choices will indeed impact on information infrastructure consistencies and dependencies, with a recurring tradeoff between tractability -brought by formal models- and expressiveness -brought by preserving contextual specificities- as well as between allowing (and maintaining) redundant data or not. Data providers have to make these choices, sometimes without a clear vision of the relative importance of end user requirements and of IT restrictions and good practices.

To address this challenge, Geonovum have started a co-design process with IT specialists to adapt UML design patterns to the requirements of an INSPIRE data extension, using the realization pattern. The JRC presented a register of design that will make it easier for members to share design choices. Last, the INSPIRE Data Specification Extension project aims at hiding the complexity of design patterns from data providers with a computer assisted design solution for schemas. During the project it was additionally noted that data providers should be aware of the fast advances in technology which may be relevant for information infrastructures: computer science technologies to deal with semantics on the Web (Linked data) or with textual data (Natural language processing, geo-tagging).

7 Knowledge and community management

The tradeoff mentioned above -unified formal models vs contextual specificities- is a question of knowledge management: what knowledge needs to be produced, processed, by machines or by people, in which context. Most presentations advocated a focus on context – the real world context - to tune the design and better consider people in information infrastructure; semantics should be human readable -at least for data production- and it is impossible to pretend there are common semantics throughout Europe. Indeed, context can account for most heterogeneities: background domain -as biodiversity community-, scale, cultures and languages (even within a same country as in Belgium).

For Geonovum, working on digital text is a good opportunity to progress understanding of semantics and conceptual models required to underpin various laws. Paul Janssen presented an information system relying on linked data technologies to articulate data paradigms and text paradigms. It integrates specific functionalities related to spatial zoning.

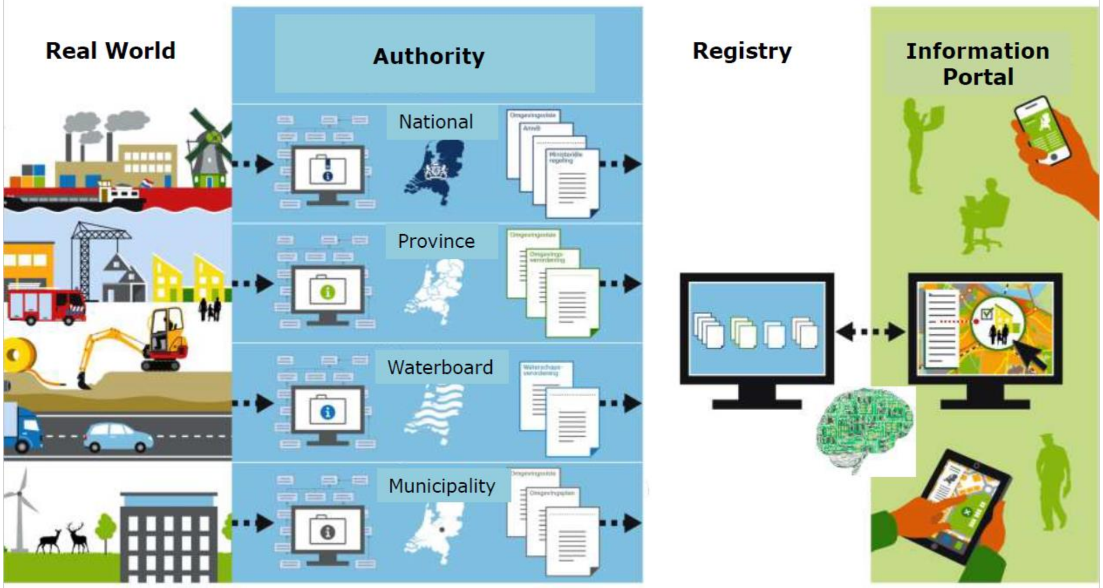

Figure 8: Integrating text document and spatial information © Janssen

8 Conclusion

The workshop evidenced that such exchange of experience and knowledge in spatial data modelling was important not only to meet European challenges but also to meet more specific national concerns. Issues encountered during reusing national data in a European information infrastructure are similar than those encountered during the design of national data portals.

Knowledge management is a recurring issue in the design of usable and scalable spatial information infrastructures. With respect to knowledge exchange, the INSPIRE Data Specification Extension project conducted by WeTransform is a promising approach to consolidate experiences and results and make them visible to the communities. The mediation approach of EuroGeographics ELS project is also a key contribution to knowledge management. Firstly it facilitates consistent browsing of data across different levels of details, which is very often essential to grab a phenomenon, its context and its composition. Second, it provides a unique entry point to the information infrastructure whereas the discovery and invocation of INSPIRE data services across different providers so far is still a tedious process. Last it connects to pan European products to reach a full coverage even on areas where the INSPIRE data services are not available. Besides, a most interesting perspective brought during the workshop was the work on smart text to improve information exchange and collaborative decision making and commitments between people, enabled by underlying digital infrastructures.

A recurring challenge was so to say the marketing of INSPIRE, the visibility of value chains. The lack of consistency between data required to monitor national and European environmental policies was pointed as a strong obstacle. One benefit of reusing high quality and documented data in an infrastructure should be in theory lowering the information uncertainty of the overall infrastructure. Yet the evaluation and documentation of information uncertainty is often performed at the level of products and by quoting the worst quality of the input sources.

In terms of investments and design, there is an interdependence of all aspects of a spatial data infrastructure in INSPIRE: technologies and standards, governance, adoption by user and providers communities. As evidenced during the presentations, obstacles often relate to lack of supporting technologies. Technologies can also create silos. Even though in theory INSPIRE and Copernicus are complementary and call for synergy, they sometimes look like two technology silos without enough true dialogue. Taken as a whole, the effort to reach a usable information infrastructure cannot be put only on the data level but also on metadata and software that process the data from different sources to inform decision-making. The objectives of the Semantic Web, or the Web of data, are quite similar to INSPIRE: data should be produced once and reused together. Technologies promoted by the Web community can contribute to INSPIRE and in particular to modelling issues encountered by participants. In this domain, EuroSDR proposed an approach consisting of designing a 'real use case to challenge scientists and software developers' to propose a solution to feed a European application with different input data sources and possibly yield a patchwork result making the most information from the input sources and uncertainties, associated with adequate documentation for result outputs.

Acknowledgement

The authors wish to thank Martijn Rijsdijk from Dutch Kadaster and Claire Ellul from University College London for their helpful comments and suggestions on this paper.

References

Video 1: https://youtu.be/xew6qI-6wNk

CDDA, 2018, https://svn.eionet.europa.eu/repositories/INSPIREDataModels/cdda-data-model/

European Parliament, Council of the European Union, 2007, Directive 2007/2/EC establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), OJL 108, 25.4.2007, pp.1-14, http://data.europa.eu/eli/dir/2007/2/oj

Grothe, Blens, Reitz, 2016, Study on INSPIRE Extensions, oral presentation at MIG-T meeting, Ispra

INSPIRE Drafting Team "Data Specifications", INSPIRE Spatial Data Interest Communities & Legally Mandated Organisations, INSPIRE Consolidation Teams and other Drafting Teams, 2013, D2.5: Generic Conceptual Model, Version 3.4rc3, http://inspire.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4rc3.pdf

Jakobsson, A., 2012, European Location Framework White Paper v1.0, http://www.elfproject.eu/sites/default/files/ELF%20White%20Paper.pdf

Masser, I., and Crompvoets, J., 2015. Building European Spatial Data Infrastructures. Esri Press, Redlands, CA, USA, 100 pgs.