What Technology does INSPIRE need?
–
Development and Research Requirements

Lars Bernard, Alessandro Annoni, Ioannis Kanellopoulos, Michel Millot, Javier Nogueras Iso, Joanna Nowak, Katalin Toth
Joint Research Centre, Institute for Environment and Sustainability
I-21020 Ispra, IT
{lars.bernard|alessandro.annoni|ioannis.kanellopoulos|michel.millot|javier.nogueras |joanna.nowak|katalin.toth}@jrc.it

SUMMARY

The proposal on an European Directive on Infrastructure for Spatial Information in Europe (INSPIRE) asks for the implementing for the components of an European Spatial Data Infrastructure (ESDI). This submission tries to link the available work for interoperable Geoinformation Service within the needs for INSPIRE. Identified technical challenges and research issues will be discussed.

KEYWORDS: *European Spatial Data Infrastructure (ESDI), Infrastructure for Spatial Information in Europe (INSPIRE).*

INTRODUCTION

The adoption of the INSPIRE proposal for a Directive (EC 2004) by the European Commission in July 2004 marks an important step on the way forward to a European-wide legislative framework that helps in achieving an European Spatial Data Infrastructure (ESDI). The proposal lays down general rules for the establishment of an ESDI for the purposes of environmental policies and policies or activities which may have a direct or indirect impact on the environment. It does not only address policy related issues but also dedicates three chapters to the technical requirements that have to be fulfilled by the member states to establish the ESDI. These three chapters are on *Metadata*, *Interoperability of spatial data sets and services*, and *Network services*. Under these chapters the proposal list general requirements on these issues as well as it formulates the requirement to adopt appropriate *Implementing Rules*. The implementing rules are meant to guide in more detail the Member States in implementing the various INSPIRE measures and to guarantee interoperability within the upcoming ESDI.

For the further INSPIRE process three phases are foreseen: a preparatory phase, from now to the adoption of the Directive that should include the drafting of the implementing rules, a transposition phase, from then until the implementation of the Directive into national legislation, an implementation phase, allowing the whole framework being made operational.

On February 2005, the Commission published the INSPIRE work programme[1] for the INSPIRE preparatory phase (2005-06). This work programme also provides an overview on an organisational and a process model for INSPIRE. The concept of Spatial Data Interest Communities (SDIC) has been introduced to provide stakeholders the mechanism to participate in the development of the implementing rules. The establishment of draft Implementing Rules will be set up in 3 different phases:

- Association phase. In this phase Drafting Teams will be created and available reference material will be collected as input for the drafting;

[1] See http://inspire.jrc.it

- Drafting phase. In this phase the Drafting Teams will establish the draft Implementing Rules;

- Review phase. In this phase a review mechanism will be set up to include stakeholder feedback through the SDICs, as well as implementation feasibility feedback from the Legally Mandated Organisations, who will be responsible for implementation of INSPIRE in the public authorities in the Member States. Finally a public consultation will be organised.

Moreover the working program hints to on-going and upcoming research projects that could provide input material to the INSPIRE drafting process and as well serve as pilots to prove the feasibility of the draft INSPIRE implementing rules.

This paper will discuss – without being exhaustive – the research issues that are identified for the scientific support of the drafting process focussing on the technical challenges in the topics harmonised data specifications, metadata, and network services and geoportals. As INSPIRE clearly also covers a wide number of socio-economic aspects, see e.g. Craglia et al. (2003) – which are not within the scope of this paper – future work will also have to address related socio-economic research issues.

HARMONISED DATA SPECIFICATIONS

According to the proposed Directive the interoperability of spatial services will be supported by the adoption of implementing rules laying down harmonised data specifications. As stated in Articles 12 and 13 the data specifications should cover

- the definition and classification of spatial objects

- the way in which those spatial data are represented

- the common system of identifiers

- and the relationship between spatial objects and key attributes

The first question to decide is how much the INSPIRE data specification process should be application driven. Three of the above mentioned tasks are within the field of conceptual modelling. The Annexes of the proposed Directive list not only the data themes for which harmonisation should be achieved, but in many cases also names concrete features (rivers, lakes, etc.) or attributes (current, salinity, wave heights of oceanographic geographical features, etc.). However only in some cases we can clearly identify current European users' needs from a European Policy. Taking *hydrography* as an example, we can derive users' requirements through the Water Framework Directive. The involvement of the SDICs in the drafting process of the implementation rules should thus bring the expertise into the process that helps to specify content, which is relevant to the European requirements or relevant cross-border applications. *The more - the better* approach is not feasible here; as point 7.2 of the Explanatory Memorandum to the Directive states that INSPIRE will be "built upon the variety of the existing information systems already in place in the Member States". A questionnaire set up in the context of the Eurospec initiative by Eurogeographics, as well as the preliminary studies conducted within the ESDI Action of JRC show that there is a clear lack of a common understanding of the terms that depict the real world's phenomena, even within a single thematic community. Clearly this problem especially arises in the case of cross-discipline harmonisation.

The results of the above mentioned initiatives evidently prove that there are many different data models both on the Member State as well as in the European level, e.g. in the fields of reference data, thematic data, etc. Thus, the questions arise

- Whether it is feasible to achieve a common European conceptual data model from all the models of the various domains?

- How to deliver translation mechanism between the various models?

- How to deal with the different quality; accuracy, or classification of data?

Parallel to the definition of the objects to be modelled the level of details should be determined as well, which on its turn determines both the spatial representation and the applicable data collection technology. Amongst others, there are three metadata elements that are widely used in qualitative description of spatial datasets: the scale, resolution, and the accuracy. *Resolution* seems to work as a meaningful qualifier in case of raster (gridded) data; but which is its equivalent in case of vector (feature) datasets? The same uncertainty is observed in case of accuracy; how can the accuracy of digital datasets be measured when data is not collected by direct measurements but indirect by digitisation of paper maps or orthoimagery? There are no standards that clearly establish the interrelationship of these properties, although they address crucial questions both for users and data producers. Information about scale, resolution, and accuracy is not only descriptive metadata, but important a priori knowledge for the conceptualisation and the spatial representation. Starting points to assure quality in the distributed INSPIRE environment are means that allow for comparison of data quality from different sources as well clear statements on the quality requirements. This enables data providers to judge the fitness of their datasets for the INSPIRE purposes, and provides guidance for their further activities.

Apart from the wide-known UML (Unified Modeling Language) there are several more or less formal approaches to describe conceptual models for geographic information. The most prominent example is the series of ISO 191xx standards for a general feature model and the Geographic Markup Language (GML 2004) for the encoding within its applications. Other examples are INTERLIS[2], which can serve as a national example implemented in Switzerland (Keller 1999); or GeoUML, an example of a regional solution (Belussi et al. 2004). INSPIRE requires such conceptual modelling approaches and notations that satisfy the special requirements posed by geographic information. Thus, research is needed to specify whether these existing tools for conceptual modelling (including the ISO standards) are sufficient for the open and interoperable encoding of geographic information.

INSPIRE Data Harmonization embraces also the issues of sharing semantics. In a distributed environment, which consists of a large number of independently developed geospatial databases, different applications have different world views, different representations, different schemas and hence different semantics. The very basic semantic interoperability can be achieved by providing metadata about the related data sets, however, to support sharing and reuse of formally represented knowledge among different systems, the concept of ontology was developed. Gruber defines ontology as a description of the concepts and relationships that can exist for a given application or discipline (Gruber, 1993). How to evolve appropriate ontologies? Should we follow the ontology hierarchy building schema (high-level ontology, domain ontology and task ontology) for every data theme listed in annexes of the INSPIRE directive or shoud we try to integrate consistent task-specific semantics to the global domain ontology? Additionally, the relationship between a meta-information model and ontologies need to be specified.

The ability of the ESDI to provide a flexible mechanism to seamlessly exchange spatial-referenced data is strongly connected to the need of a coding system that provides unique identifiers (UIDs) for the objects exchanged in the distributed ESDI environment. Existing solutions can be roughly distinguished in approaches using human-readable or machine generated UIDs. Further research

[2] http://www.interlis.ch/index_e.htm

should help to analyse the requirements on UIDs as well as existing solutions and give guidance on how to design a UID system that is feasible within the ESDI.

The European data harmonisation process should be inclusive. Due to the different development levels of digital spatial data production it seems to be reasonable to set up specification families that share the conceptual model, but differ in the spatial representation and the degree of generalisation. This multiple levels of representation require that questions of model and geometric generalisation should be equally addressed. In model generalisation a logical and consistent method is necessary where the objects of the smaller scale representation consequently encapsulate the objects of the bigger scale representation. Moreover the function for the unambiguous mapping of one object into another one should be determined together with criteria for the inclusion of an object in the generalisation chain.

Considered from a geometric perspective those generalisation/simplification methods are efficient that:

- guarantee the integrity - topological consistency - at all resolution levels,
- produce predictable and repeatable results,
- reduce at maximum the data volume for the chosen resolution,
- minimise the deviation of features from their original position,

Some methods seem to work quite efficiently in an automatic way in case of certain datasets (for example the Duda-Hart method in case of the highly hierarchic administrative boundary databases), but more complex datasets require more human interaction. The comparative analyses and testing of the generalisation methods together with predicting the necessary preliminary processing is one of the most challenging technological tasks for the implementation of INSPIRE.

METADATA

Geospatial data catalogues can be considered the heart of any SDI as they provide the means to publish descriptions of geospatial data holdings and services in a standard way to permit search across multiple servers. These descriptions, known as metadata, are the information and documentation which makes data understandable and sharable for users and applications over time. Thus, any improvement in the management, creation and understanding of metadata will represent a general improvement for SDI interoperability.

The INSPIRE proposal requires the Member States to provide descriptions of the available spatial data sets and services in the form of metadata. Furthermore it requires that these metadata are properly maintained and kept up to date. Some of the first tasks that according to the proposal need to be accomplished include the identification of the elements to be included in the metadata documentation, following also the INSPIRE proposal requirements, and the appropriate standards that need to be followed in this process.

The use of ontologies has an important role also in respect to metadata, as it will facilitate the semantic interoperability between metadata created in different contexts and by different communities. For instance, the results of the Simple Knowledge Organization System (SKOS) activity of the SWAD-Europe project[3] show how to use ontology modelling, encoding and the establishment of the connections among different ontologies.

[3] http://www.w3.org/2001/sw/Europe/

Multilingual aspects of metadata should be considered within the European context. Member states are not expected to provide translation for each metadata record they produce. However, a European SDI catalog must tackle the problem of finding resources independently of the language used for metadata creation. Automatic translation technologies, multilingual gazetteers (many user queries are mainly based on geographic locations) and again the use of multilingual ontologies should facilitate this labour.

Existing European metadata bases use different standards and schemas. With respect to the geographic context, ISO19115 metadata standard seems to unify all previous metadata standardization initiatives. However, different profiles of this standard will be needed and it is also important to provide visibility to other communities, e.g., providing translations to more general standards as the Dublin Core standard. Therefore, the construction of crosswalks and mappings between standards is an important issue that must be considered, not only to solve syntactic heterogeneities but also to solve semantic heterogeneities. Clearly the latter is tightly coupled to the tasks on semantic interoperability for harmonised data specifications as described in the chapter above.

The relationships between descriptions of different resources should become transparent. For instance, metadata describing feature instances should be connected to metadata description of feature types, the datasets compiling all these feature instances, or even dataset series compiling previous datasets. Descriptions of all these related items should be consistent, coherent and non-redundant. In order to establish these relations, universal identification of resources (both data and metadata) plays a fundamental role that must be faced as soon as possible (see also previous chapter).

Finally, the SDI community should also consider seriously how to integrate better metadata creation within the general process of data creation. For instance, every time a dataset is modified, the data provider should be committed to include a new process step in the lineage contained within the data quality section of metadata. Training, methodologies and policies for data creation should be revised to include metadata creation as a necessary and fundamental step within the general data creation workflow. Furthermore development of metadata tools which will automate as much as possible the process of metadata creation and maintenance is necessary and will facilitate this process.

NETWORK SERVICES AND GEOPORTALS

The INSPIRE proposal distinguishes five service types:

- Discovery services
- Data view services
- Download services
- Transformation services
- "invoke spatial data services" services

Following the INSPIRE proposal, the network services will be established in a distributed environment (having at least one service access point per Member State) and they will be accessible via the European Geo-Portal. The definition of appropriate technical specifications requires that considered interface specifications are mature and proved by implementations and operational usage. The first tasks in this will have to provide a more detailed description as a basis for the common understanding about these network services. This also requires defining a software architecture that describes the collaboration of the network services and the Geo-Portal addressing for instance

- General architectural model (how to invoke network services, functionalities of the EU Geoportals, …)

- Security (access to the service and data transfer)

- Multilingualism

- Conformance

- Performance

- Confidentiality

- Technical architectures and protocols

Various SDI implementations prove that available technical specifications for interoperable geoinformation services provide a good basis for operational usage. However the gained experiences also show that this set of standards and specifications is not yet complete and fully mature (Bernard et al. 2005). The areas where further development and research will most certainly be welcomed can be split into 2 categories:

- Towards operation

- Towards Enhanced Capabilities

Towards Operation

Once the directive is approved, the setting of the INSPIRE will lead to a significant activity in the European Union Member States. This unprecedented activity geared towards the development of interoperable services will pose numerous challenges and in this paper, we stress the implementation and integration issues.

The implementation in parallel and in the whole European Union may require to further develop existing standards to alleviate any ambiguity and to provide the necessary additional levels of details for implementation. Any ambiguity or lack of details may lie with the functionality themselves that is tackled early in the drafting of the implementing rules, but they may as well be with non functional requirements not fully addressed such as the operational behaviour of the services, for example:

- *capacity*, (how much of a capability is needed at any moment in time), for example how much of the Web Features service transaction request is needed ?.

- *speed* (states how fast the complete operation, or sequence of operations, is to be performed) that clearly have an impact on the implementing rules should a miminal performance requirement be requested for a given service. For example should a member state discovery service be distributed, the temporary non availability of one catalogues service may have to be explicitly considered and potentially not only through the maximum number of returned records.

Other implementation constraints that may require further development of the candidates for the technical specifications are:

- *Adaptability* that measures how easily a system copes with requirements changes that may be necessary as the user requirements are modified by experience, taking into account for example the diverse situation in the Member states or the evolution of the reference standard.

- *Security*: It is envisioned that the INSPIRE services may need to be secured against threats to confidentiality, integrity and availability, in addition the provision of some services may require the exchange of confidential information.

- *Safety* such as consequences of software failure should be made clear for example in the critical case of a Web Feature Service update transaction operation.

- Detailed *conformance* testing that may require the development of one or several comprehensive conformance testing environments.

- Communication interface where *performance* attributes may be associated to some elements of the service specification such as networks and networks protocols.

For services without an existing off-the-shelf proposal, development and research will be required to reach the state where the above consideration can be efficiently addressed. Example of such services is the transformation and the "invoke spatial service" services.

The integration of the INSPIRE network services with other supportive services already identified in the directive proposal such as e-commerce services or access control services may drive the further development of the existing technical solutions that do not yet formally and explicitly address access control, secure payment or digital rights management.

From an architecture perspective, the INSPIRE directive proposal asks for the definition of individual services, but European Union Member States may provide access to these services through their own portal so it may also be necessary to further develop existing standalone specifications to take into account the architectural considerations of geo-portals.

Finally the INSPIRE services will most certainly cohabit with existing services and infrastructures, an aspect that existing specifications may not have taken into account. This generic constraint may then drive additional adaptation of existing standards

Towards Enhanced Capabilities

In addition to short to medium term considerations linked with the operational set-up of the INSPIRE network services, it is possible to envision some enhancements in relevant research fields that could prove beneficial and we can take a look at them from the service perspective or the application perspective.

Under the assumption that INSPIRE services will be web services the maturation and stabilization of the applicable standards or concepts such as the ones related to the Service oriented Architecture can only prove beneficial. In the European context another short term requirement here clearly links to the lack of multilingualism support within existing service specifications. On a longer term perspective the operational introduction of semantic in services, such as discovery services or download and transformation services, coupled with intelligent agents technology could lead to a new and exiting dimension of spatial data services.

On the efficient use of existing services and or resources, new development and maturation of existing development in service chaining and workflow management engines or grid technology could increase the benefits one can have from these services.

BIBLIOGRAPHY

Belussi, A., Negri, M., Pelagatti, G., 2004. GEOUML: A Geographic Conceptual Model Defined Through Specialization Of Iso Tc211 Standards. In: Proceedings of 10th EC GI & GIS Workshop, ESDI State of the Art, Warsaw, Poland, 23-25 June

Bernard, L., Kanellopoulos, I., Annoni, A., Smits, P., 2005. The European Geoportal - One step towards the Establishment of a European Spatial Data Infrastructure. Computers, Environment and Urban Systems, 29: 15-31.

Craglia, M., Annoni, A., Klopfer, M., Corbin, C., Pichler G., Smits, P., 2003. Geographic Information in the Wider Europe - GINIE. Final Report. European Commission.

EC, 2004. Proposal for a DIRECTIVE of the European Parliament and the Council of the European Union of 23 July 2004 on establishing an infrastructure for spatial information in the Community (INSPIRE).

GML, 2004. Geography Markup Language. Version 3.1 Public Release from Open GIS Consortium, CoverPAGES, March 26, 2004. http://xml.coverpages.org/ni2004-03-26-a.html

Gruber, T., 1993. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal Human-Computer Studies 43, p.907-928.

Keller, S.F, 1999. Modeling and Sharing Graphic Representations of Geographic Data with INTERLIS. In: Proceedings of interop'99, Zurich.