ISO Workshop on address standards: Considering the issues related to an international address standard
25 May 2008, Copenhagen, Denmark, ISBN 978-1-86854-689-3

Addressing the needs of INSPIRE: The Challenges of improving Interoperability within the European Union

Andrew Coote

Facilitator, Addresses Thematic Working Group
andrew.coote@consultingwhere.com

Disclaimer

The views expressed in this paper are personal and do not necessarily reflect those of the INSPIRE coordination team or the European Commission itself. Text used within INSPIRE documents has been extensively used and must not be reproduced without their express permission.

1. Background

Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 established an Infrastructure for Spatial Information in the European Community (INSPIRE) and it entered into force on 15 May 2007.

The purpose of such an infrastructure is to assist policy-making in relation to policies that may have a direct or indirect impact on the environment. INSPIRE is to be based on the infrastructures for spatial information that are created by the Member States and are designed to ensure that spatial data are stored, made available and maintained at the most appropriate level so that it is possible to combine spatial data from different sources across the Community in a consistent way and share them between multiple users and applications.

The Directive focuses in particular on five key areas:

- metadata;
- the interoperability and harmonisation of spatial data and services for selected themes (as described in Annexes I, II, III of the Directive);
- network services and technologies;
- measures on sharing spatial data and services;
- coordination and monitoring measures.

This paper focuses upon the work to improve interoperability and harmonisation through creation of a neutral data specification for Addresses. The significance of the inclusion of Addresses in the Annex I themes of the Directive is that these themes have been identified as those with the greatest significance to environmental applications and therefore prioritised for action. Annex I themes also includes cadastral parcels, administrative units, geographical names, hydrography, protected places and their underlying coordinate reference systems. In the context of Addresses, this provides an exciting opportunity to influence data specifications for most, if not all, the key object types that addresses reference.

A lot of ground work has already been completed. During the preparation of INSPIRE proposals, an international team of experts was formed to review available reference material and international standards to create a set of prime reference documents that govern the development of the data specifications. This "conceptual framework" consists of, amongst others, the following key documents:

- D2.3 Definition and scoping of the Annex themes;

- D2.5 Generic Conceptual Model;
- D2.6 Methodology for the development of data specifications;
- D2.7 Guidelines for the encoding of spatial data.

Consequently, at the time of formation of the Thematic Working Group (TWG) on Addresses in February 2008, we are in a fortunate position in that a vast array of background material had been assembled, a methodology for data specification developed and ground rules for interoperability available.

2. Terms of Reference

Our different roles in analysing, drafting and reviewing can be summarised as follows.

- to analyse the reference materials and the user requirements submitted by the INSPIRE stakeholders for the specific spatial data theme;
- to analyse and take into account the standards developed by organisations established under the international law;
- to seek advise of the EIONET (European panel of environmental experts) at the initial identification of requirements and any time when there is a need to do so;
- to apply the INSPIRE conceptual framework to ensure conformance to the INSPIRE Generic Conceptual Model and overall consistency in specification development process applying the agreed methodology;
- to apply, whenever possible, the terminology, coordinate reference system, the feature catalogue, and other items maintained in the INSPIRE registers;
- to propose specified items or changes to existing items to the maintainers of the registers;
- to ensure consistency, with other themes;
- to document the specification process with full references and justifications,
- to analyse and consolidate the comments received from the Commission, SDICs and LMO as a result of document reviews and practical testing of the specification;

3. Deliverables

The TWG is tasked with delivering the following:

- Analysis of the reference materials;
- Analysis of the user requirements and documentation of the use-case development;

- Data specifications covering the prescribed aspects (discussed in more detail later in this paper);

- Resolutions of the comments submitted by the stakeholders in frame of reviewing the draft data specifications;

- Provide material that can be used to communicate the work of INSPIRE.

4. TWG Team

The team initially appointed, as a result of selection from applications to the Commission in early 2008, was somewhat skewed towards experts from Western and Northern Europe. This was recognised at our first meeting as a serious weakness. Ownership by all stakeholders (countries) in any project of this nature is a critical success factor. Subsequently we have been able to redress this balance by the addition of a member from a relatively recent accession country, the Czech Republic, and another from Southern Europe. The full team of address experts is now:

Alicia Gonzalez	Spain	Jan Zindulka	Czech Republic
Morten Lind	Denmark	Nick Griffith	UK
Per Sundburg	Sweden	Sara Greenwood	UK
Udo Maack	Germany	Yvette Ellenkamp	Netherlands
Ziggy Vanlishout	Belgium		

There are three additional members of the group:

Darja Litheneger	Representing the INSPIRE coordination team within the Commission;
Andrew Coote	Facilitator – chair of the group;
Andres Friis-Christiansen	Editor – responsible to translating the work of the group into the prescribed format, particularly in relation to the Unified Modelling Language.

5. Methodology

A well defined methodology has been set out for our work, involving a series of steps as follows:

- User requirements are described in use cases and application scenarios. These use cases will typically be described according to a template for use case descriptions. The result is a description of the relevant universe of discourse.

- An "as-is" analysis of the current situation is carried out in parallel, identifying the relevant data harmonisation components (see next section).

- The gap analysis identifies user requirements that cannot be met by the current data offerings. For each gap, a data harmonisation approach - which may also include a conclusion that specific user requirements cannot be met - needs to be identified and agreed upon.

- This approach is then documented as an application schema describing required spatial object types with their constraints, properties and range of property values. This establishes an agreed terminology. The application schema is described in a conceptual schema language. The Unified Modelling Language (UML) as profiled by the ISO 19100 series of International Standards will be used as the conceptual schema language. Together with the corresponding feature catalogue, this application schema constitutes a core component of the Data Specification.

- The Data Specification will be documented according to ISO 19131, the International Standard specifying the contents of Data Product Specifications in the field of geographic information. Data Product Specification includes at least the following sections: specification scope, data product identification, data content and structure, reference systems, data quality, data product delivery, and metadata. If possible additional items, such as maintenance, portrayal, data capture may also be included.

- Based upon the application schema, encodings for the delivery of data according to the data specification will be specified. It is expected that where appropriate a GML (Geography Markup Language, ISO 19136) application schema will be generated following the normative rules for such conversions. Additional delivery formats, for example for data with coverage functions, will be defined based on specific user requirements and the state-of-the-art in the relevant communities.

- These results will be tested within a pilot under real world conditions.

- Incremental costs and benefits of the harmonisation efforts will be tracked and documented.

6. Data Harmonisation Components

This is the term used to describe the key areas of the data specification that have been identified as likely to require consideration during the project. The following is a summary of the components identified:

(A) Principles	The principles of the Directive are considered to be a general basis for developing the data harmonisation needs. The first three of the five principles are to be considered to help define the data harmonisation process: - *that spatial data are stored, made available and maintained at the most appropriate level;* - *that it is possible to combine spatial data from different sources across the Community in a consistent way and share them between several users and applications;* - *that it is possible for spatial data collected at one level of public authority to be shared between other public authorities.*
(B) Terminology	This component will support the use of a consistent language when referring to terms via a glossary. This needs to be registered and managed through change control with multi-lingual support.

(C) Reference model	Define the framework of the technical parts including topics like information modelling (i.e. conceptual modelling framework with rules for application schemas) and data administration (i.e. reference systems).
(D) Rules for application schemas and feature catalogues	The purpose of this component is to - provide a computer-readable data description defining the data structure - enabling automated mechanisms for data management; - achieve a common and correct understanding of the data, by documenting the data content of the particular theme, thereby making it possible to unambiguously retrieve information from the data. Feature catalogues define the types of spatial objects and their properties (attributes, association roles, operations) as well as constraints and are indispensable to turning the data into usable information. Feature catalogues promote the dissemination, sharing, and use of geographic data through providing a better understanding of the content and meaning of the data. The full description of the contents and structure of a spatial dataset is given by the application schema which is expressed in a formal conceptual schema language. The feature catalogue defines the meaning of the spatial object types and their properties specified in the application schema. Text elements in the feature catalogues should be maintained at least in the official European languages
(E) Spatial and temporal aspects	Conceptual schema for describing the spatial and temporal characteristics of spatial objects: - Spatial geometry and topology - Temporal geometry and topology - Coverages (examples of coverages include rasters, triangulated irregular networks, point coverages, and polygon coverages) While the component "reference model" specifies an overall framework, this component deals with the spatial and temporal aspects in more detail, for example, the types of spatial or temporal geometry that may be used to describe the spatial and temporal characteristics of a spatial object.

(G) Coordinate referencing and units of measurement model	This component will describe methods for spatial and temporal reference systems as well as units of measurements – including the parameters of transformations and conversions. The focus is on reference systems that are valid across Europe (in case of projected systems split into zones this will be a collection of such systems covering the different zones). This component will also support European geographical grids.
(F) Multi-lingual text and cultural adaptability	Conceptual schema for multi-lingual character strings in spatial objects and supporting information: - To be used in all application schemas and as a result in data instances: all string valued properties that may be provided in a language shall use this type - To be used in the dictionary model so that dictionaries may be multi-lingual, e.g. the feature catalogue, the feature concept dictionary or codelists. At the moment, it is not planned to document the formal application schema (classes, attributes, associations, constraints) in multiple languages as the definitions are part of the feature catalogue and data dictionary. In principle, cultural differences have to be taken into account, e.g. not all terms may be translatable from one language to another. Furthermore, cultural differences between communities working in the same language can be at least as much a problem as multi-lingual issues. Ontologies may help to capture multi-cultural aspects.
(H) Object referencing modelling	This component will describe how information is referenced to existing objects, typically base topographic objects, rather than directly via coordinates. It will be specified how the spatial characteristics of a spatial object can be based on already existing spatial objects. As a result, this component will support the generation and maintenance of application-specific “user geographies” based on reference data.

(I) Data translation model / guidelines	This component is about translating from a national/local application schema to the INSPIRE application schema and vice versa. Translations are required for data and for queries. NOTE No well-defined set of translation capabilities has been standardised in the GI community at this time. It is not yet clear, if there will be a need to specify translations also between different European application schemas, e.g. for different representations or for creating specific information products, e.g. for reports or from base data etc. Also, further research would be required to identify how consistent adoption of ontologies could be exploited here.
(J) Portrayal model	This component will define a model for portrayal rules for data according to a data specification. It will clarify how standardised portrayal catalogues can be used to harmonise the portrayal of data.
(K) Identifier management	Spatial objects should have an external object identifier. This component will define the role and nature of unique object identifiers (or other mechanisms) to support unambiguous object identification. To ensure uniqueness some form of management system will be required. This does not mean that all organisations need to adopt a common form of identifier or other mechanism but the identifier management mechanisms (e.g. registers) in use at national level will need to be synchronised / mapped to ensure pan-European integration. Note that the same real-world phenomenon may be represented by different spatial objects (with their own identifiers).
(L) Registers and Registries	Registers will at least be required for : - all reference systems used in spatial data sets - all units of measurement used in spatial data sets - all codelists / thesauri used in the application schemas (multi-lingual, at least in all official European languages) - the feature concept dictionary for elements used by application schemas (multi-lingual, at least in all official European languages) - identifier namespaces - all feature catalogues - all application schemas

	The registries will be available through registry services. Metadata on dataset level will be available through catalogue services.
(M) Metadata	Metadata is to be included on the following levels: - Discovery - Evaluation - Use Metadata associated with individual spatial objects will in general be described as part of the application schemas.
(N) Maintenance	Best practice in ensuring that application data can be managed against updates of reference information without interruption of services. This will require the definition of mechanisms by different stakeholder areas to manage where this is required and it is feasible: - change only updates - versioning of objects (and their properties) - object lifecycles Propagation of changes across scale and between dependent objects is required in general to maintain consistency of the data (automatic or manual processes).
(O) Data & information quality	This component will advise the need to publish quality levels of each spatial dataset using the criteria defined in the ISO 19100 series of standards, including completeness, consistency, currency and accuracy. This will include methods of best practice in publishing: - Acceptable quality levels of each spatial dataset - Attainment against those levels for each spatial dataset Quality information associated with individual spatial objects is part of the metadata associated with the respective spatial object (see component "Metadata") and will in general be described as part of the application schemas.
(P) Data transfer	This component will describe methods for encoding application and reference data as well as information products.

	The encoding of spatial objects will in general be model-driven, i.e. fully determined by the application schema in UML. To support network services that are implemented as web services, spatial objects are expected to be primarily encoded in XML/GML for the exchange of spatial data. Coverage data is expected to use existing encodings for the range part.
(Q) Consistency between data	Guidelines are required as to how the consistency between the representation of the same entity in different spatial datasets (for example along or across borders, themes, sectors or at different resolutions) will be maintained. The custodians of such spatial datasets will decide by mutual consent on the depiction and position of such common spatial objects or they will agree on a general method for edge-matching or other automatic means to maintain data consistency.
(R) Multiple representations	This component will describe best practices how data can be aggregated: - across time and space - across different resolutions ("generalisation" of data)
(S) Data capturing rules	Here we will describe the data specification-specific criteria regarding which spatial objects are to be captured and which locations/points will captured to represent the given spatial object (e.g. all lakes larger than 2 ha, all roads of the Trans European Road Network, etc.). For INSPIRE data specifications it is in general not relevant how the data is captured by the data providers.
(T) Conformance	This will define how conformance of data to a data specification is tested, i.e. it will be necessary to apply conformance tests as specified in the individual data specification. Ideally these will be automated. In addition, all INSPIRE data specifications will conform to the Generic Conceptual Model as well as, since Data Specifications are specified using ISO 19131 (Data product specification), to ISO 19131.

7. Progress

At the time of writing the TWG has been operating for a little under three months, although a number of members have only been appointed more recently. However, we have already been able to make some substantial progress:

Team-building

We have conducted more than a dozen telephone conference calls, some to gauge progress on work items and others to allow more detailed discussions on particular topics. The team are working to a project plan developed in Microsoft Project that is regularly reviewed and revised.

Reference material

A set of reference materials has been compiled from many sources. Including standards, specifications and implementation guidelines, the materials are not only from European Union countries but also other parts of the world, and include commercial sources. The documents have all, to some level, been translated into English.

Data Description

The theme definition for addresses, has been rewritten by the group. For instance, we have redefined the scope of property to enable us to consider a wider range of objects (features in ISO terms) that might be "addressable". We have also extended the range of terms in the INSPIRE glossary to help our deliberations.

User Requirements

The results of the User Requirements Survey (undertaken by the Commission) has been somewhat disappointing, most submissions being expressed at a very general and non-application-specific level. However, we have been able to make considerable progress using experience of our own experts, supplemented by the reference material. We now have a document that lays out user requirements at an application level (expressed in terms of business functions). It also contains a more detailed system-level description of generic functions that any system accessing a database containing address data will need to fulfil. These two "views" of the requirements are cross-referenced. In addition, we have documented the actors and high-level processes in UML. The plan is now to circulate this document to potential users to validate our analysis.

Application Schema

A first cut application schema has been produced, working from the Swedish model, which is itself based on ISO 19112. Work is now in progress to compare it to each of the reference materials. The primary focus, at this stage, is to identify data elements, either object-type or attribution (including what might be referred to as object-level metadata). This will aid our discussions with the other thematic working groups that are working on areas that impact strongly on addressing, particularly administrative units, geographical names, transport networks and cadastral parcels.

Data Harmonisation Components

A sub-group formed to look at these components has produced a "position paper" setting out their views on a number of the technical priorities including terminology, temporal attribution, coordinate referencing and metadata. This will provide a platform for debate at our two day face to face meeting in Copenhagen following the ISO meeting.

Enterprise Architect

By the time this paper is presented we expect to have loaded our initial application schema and other materials into our repository tool, Enterprise Architect. This will be an important milestone since it will prove (and possibly show the limitations of) the development methodology we have adopted.

8. Key Issues

The work we have undertaken has not been without pain and has exposed a number of issues. Many of those experienced in International standards work will be familiar with some of these.

Language - the first language of many of the members of the team is not English and yet it is the language of our deliberations. It is difficult enough to get complex, detailed arguments across in your first language without the added complication of using another (supremely irregular) one.

Distributed Team – being located all over Europe, the group communicates mostly by email and teleconference. Whilst telephone conferences are extremely helpful, there is no substitute for face to face meetings when it comes to debating key conceptual issues.

Cross-theme Overlaps – the Annex 1 themes are being worked on, in parallel, by eight different working groups. At this stage, most are trying to get their ideas clear on their own scope without too much attention to the potential overlaps (and gaps). The Address working group is particularly vulnerable in this respect, since our scope intersects so many different themes.

9. Conclusions

The work of the INSPIRE Thematic Working Group is at a very early stage. However, there are many parallels between what it is doing now and what ISO may decide to embark upon in creating an International standard. In our work to date, there is a strong will between all the technical experts to reach consensus and I am confident that the result will be an extremely valuable contribution to understanding of addresses and their interoperability.

The process of then translating such a technical agreement, through a management and political debate, into a coherent Europe-wide implementation will be another challenge.