Evolution of Environmental Information Models

Katharina Schleidt

Evolution of Environmental Information Models

Katharina Schleidt

Umweltbundesamt GmbH Austria
Katharina.Schleidt (@) umweltbundesamt.at

Abstract. Access to environmental data, based on standardized data models and services, is becoming ever more prevalent, providing stakeholders with access to a wide range of standardized environmental data from diverse sources. However, exactly this success brings new problems, with thematic extensions based on these standardized models being created by disparate thematic communities based on their specific requirements. In contrast to the traditional standards development process, which includes mechanisms for maintaining alignment of concepts across different sections of the standard, once these standards are extended by a larger and not so strictly structured community, the alignment process becomes increasingly difficult. This position paper sketches this problem, as illustrated by example of the European INSPIRE process, and serves as a basis for the conference workshop discussion that aims to capture both further facets of the problem as well as possible solutions.

1   Introduction

Access to environmental data is necessary for environmental research and control; often, this requires the use of data from different sources. This in turn entails a tedious process of accessing data in various formats, identifying the relevant concepts in the various data, then aligning and merging this data before the necessary analyses can be performed.

To ease this process, various initiatives have been launched around the world with the aim of providing standardized access to environmental data. While these initiatives differ in thematic focus, spatial extent or governance, they all strive to provide standardized data models and service specifications for the access to and use of environmental data.

Based on these developments, easy access to standardized and harmonized environmental data should be a simple process. Unfortunately, this initial impression is deceptive. Many of the existing standards only cover the core concepts of a domain; extensions are left to the thematic communities using these standards. As these thematic extensions grow and develop, they rapidly introduce new concepts not aligned

with similar concepts stemming from a different thematic area [1], we find ourselves dangerously close to the starting point.

Thus, an ongoing revision process becomes necessary, with many similar characteristics as the initial standardization process, but with additional challenges such as:

- A wider participant scope, as the adoption of the existing standardized data models leads to wider uptake;
- More complex concepts, as the scope widens through extensions;
- Less governance, as officially the problem is considered to be solved.

This paper sketches the requirements for such ongoing collaboration on a standardized data model for environmental data. It gives an overview of the types of problems already identified as well as those likely to happen in the foreseeable future. Based on these findings, it describes possible mechanisms to support this process.

2 Background

Various initiatives, stemming from various environmental sub-domains, have been launched in the last years with the goal of allowing for easy access to relevant data through the standardization of data models and service specifications. These include thematically narrow initiatives such as the Long Term Ecological Research or Biodiversity communities (see the ILTER[1], GBIF[2] and TDWG[3] sites) as well as thematically broad initiatives such as Research Data Alliance[4] or the European INSPIRE Initiative [2]. While data standards are often in place for core concepts, there is always a need to extend these in order to support new or alternative requirements.

In this paper, we use the European INSPIRE Initiative [2] as an illustrative example of both the challenges faced in such an undertaking as well as to provide examples of ways to manage these problems. However, the challenges described are the same for all initiatives attempting this task, and thus the conclusions reached have a wider validity.

2.1 The INSPIRE Directive

The INSPIRE Directive entered into force on the 15th of May 2007. It recognizes that definition and enforcement of European Community policy on the environment requires the easy availability of high quality spatial data. The Infrastructure for Spatial Information in the European Community (INSPIRE) should assist policy-making in relation to policies and activities that may have a direct or indirect impact on the environment. The 34 spatial data themes covered by INSPIRE have been structured in the three annexes to the directive, the themes are as follows:

---

[1] http://www.ilternet.edu/
[2] http://www.gbif.org/
[3] http://www.tdwg.org/
[4] https://rd-alliance.org/

ANNEX I

1. Coordinate reference systems
2. Geographical grid systems
3. Geographical names
4. Administrative units
5. Addresses
6. Cadastral parcels
7. Transport networks
8. Hydrography
9. Protected sites

ANNEX II

1. Elevation
2. Land cover
3. Orthoimagery
4. Geology

ANNEX III

1. Statistical units
2. Buildings
3. Soil
4. Land use
5. Human health and safety
6. Utility and governmental services
7. Environmental monitoring facilities
8. Production and industrial facilities
9. Agricultural and aquaculture facilities
10. Population distribution — demography
11. Area management/restriction/regulation zones and reporting units
12. Natural risk zones
13. Atmospheric conditions
14. Meteorological geographical features
15. Oceanographic geographical features
16. Sea regions
17. Bio-geographical regions
18. Habitats and biotopes
19. Species distribution
20. Energy resources
21. Mineral resources

INSPIRE is based on the European Member States (MS) spatial information infrastructures, that are made compatible with common implementing rules and are supplemented with measures at Community level. These rules and measures should en-

sure that the infrastructures for spatial information created by the MS are compatible and usable in across administrative levels and national boundaries.

Due the wide diversity of formats and structures in which spatial data are organized and accessed in the Community, data specifications have been provided to facilitate the use of spatial data from different sources across the MS's. Network services for sharing spatial data between the various levels of public authority in the Community make it possible to discover, transform, view and download spatial data and to invoke spatial data and e-commerce services.

3   Process and stakeholders

Before we go into technical details, we must first consider the processes required for the creation, and subsequent extension of a harmonized data model, as well as the various stakeholders involved in this process. According to Craglia [3], INSPIRE has some characteristics that make it particularly challenging [1]:

1. The infrastructure is built on those of 27 Member States of the European Union in more than 23 languages. This requires the coexistence and collaboration of very different information systems, professional and cultural practices,
2. Given this complexity, it was necessary to adopt a consensus-building process, involving hundreds of national experts, to develop the technical specifications for INSPIRE,
3. Existing standards must be tested in real distributed and multilingual settings,
4. Standards that are not mature enough, or leave too much room for different interpretation (because of the legally mandated implementation) have to be refined,
5. Standards which do not yet exist must be developed,
6. Inconsistency and incompatibility of data and metadata must be addressed for the 34 themes that fall within the scope of the Directive [1]

In order to counter these challenges, the following development process was defined:
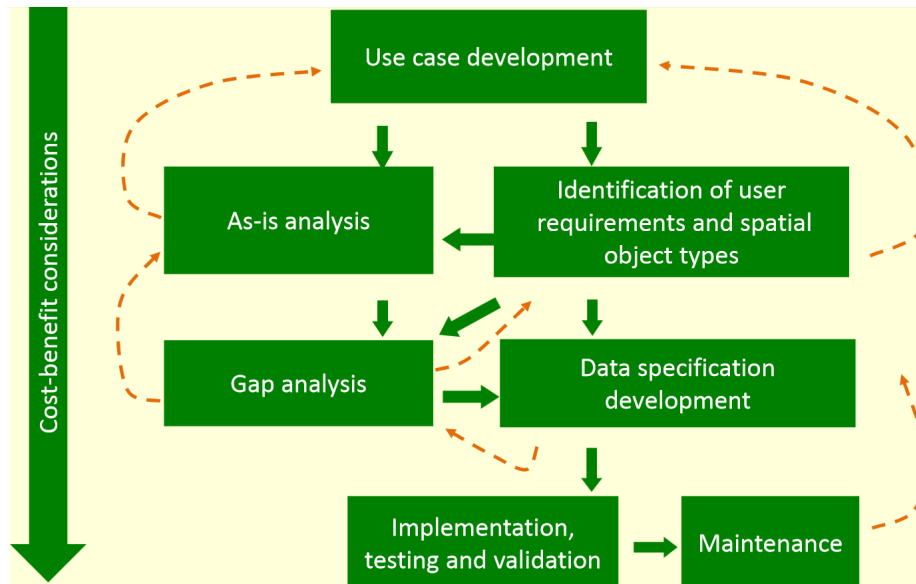
Fig. 1. Steps in the INSPIRE data specification cycle [1]

All the tasks leading to the Data specification development were undertaken by the members of the Thematic Working Group (TWG) constituted for developing the data model for a specific theme. Based on the use cases provided, and utilizing the ISO 191XX suite of standards, data models have been developed for the representation of the data themes listed above. The TWG members were supported by JRC staff in finding points for alignment between themes, as well as making the best use of existing data types stemming from both the underlying ISO standards as well as the INSPIRE base models.

Once the data specification development was advanced to the point where the TWGs were satisfied with their thematic data models, these data models together with their specification were made available to interested parties from the European Member States (MS). The MS were encouraged to test these preliminary data models both through manual scrutiny of the data specifications provided as well as by filling these data models with data stemming from their national data holdings. The feedback from this testing and validation process was in turn sent to the responsible TWGs, for a second round of data specification development (with all ensuing feedback loops illustrated above). More information on the INSPIRE data specification development process can be found in the JRC deliverable D2.6 Methodology for the development of data specifications [4]

Unfortunately, while the implementation, testing and validation step performed by the MSs did give much valuable feedback on thematic requirements, it provided little support for the task of harmonizing and aligning the data model itself. This is due to the fact that this thematic feedback usually came from dedicated thematic departments within the national environmental agencies as well as other stakeholders. Thus, while

let's say the air quality department provided thorough feedback on the requirements stemming from air quality monitoring and the water management group provided feedback on requirements stemming from the water domain, there was little consideration on the fact that similar concepts from these two domains could be encoded in the same manner; this task was left to the TWG members as well as the coordinators from the JRC. Thus, despite best effort from all parties involved, inconsistencies in the INSPIRE data models were not avoidable.

Already within the harmonized INSPIRE data specifications, despite the constraints set out by the INSPIRE Generic Conceptual Model [5], equivalent concepts from individual thematic data specifications vary due to difficulties in alignment across thematic domains.

Now that the INSPIRE data and service specifications have been finalized and are being implemented on both the national as well as the European levels, further extensions are being created of these base INSPIRE data models. For example, an extension of the INSPIRE data models for European Air Quality Reporting has been finalized and is now operational[5]; the same approach is currently being followed for various other environmental reporting obligations. While these extended data models will be tailored to the requirements of a specific reporting obligation, they will undoubtedly be adding certain similar concepts such as an European station code to the base INSPIRE models. At present there is no mechanism in place to enable alignment of equivalent concepts across the various pending extensions of the core data models.

4   Examples

In the following section, we shall provide some examples from the INSPIRE domain. In the first example, we shall show how inconsistencies crept into the tightly governed INSPIRE data specification process. In the second example, we shall show how this problem will certainly be exacerbated now that the INSPIRE data models are available and being extended for thematic purposes, with little or no governance to assure alignment.

### 4.1   INSPIRE Data Specifications - Tight Governance

In theory, the INSPIRE Generic Conceptual Model (GCM) as well as the Methodology for the development of data specifications govern the data specification process. However, despite vigilance by all parties concerned, discrepancies appeared such as follows:

- In the INSPIRE data specification for Area management/restriction/regulation zones and reporting units, the name of the zone is defined as the INSPIRE base type GeographicalName.
- In the INSPIRE data specification for Environmental Monitoring Facilities, the name of the facility is defined as the ISO 19103 type CharacterString.

---

[5] http://www.eionet.europa.eu/aqportal

Fig. 2. Example - Name in INSPIRE
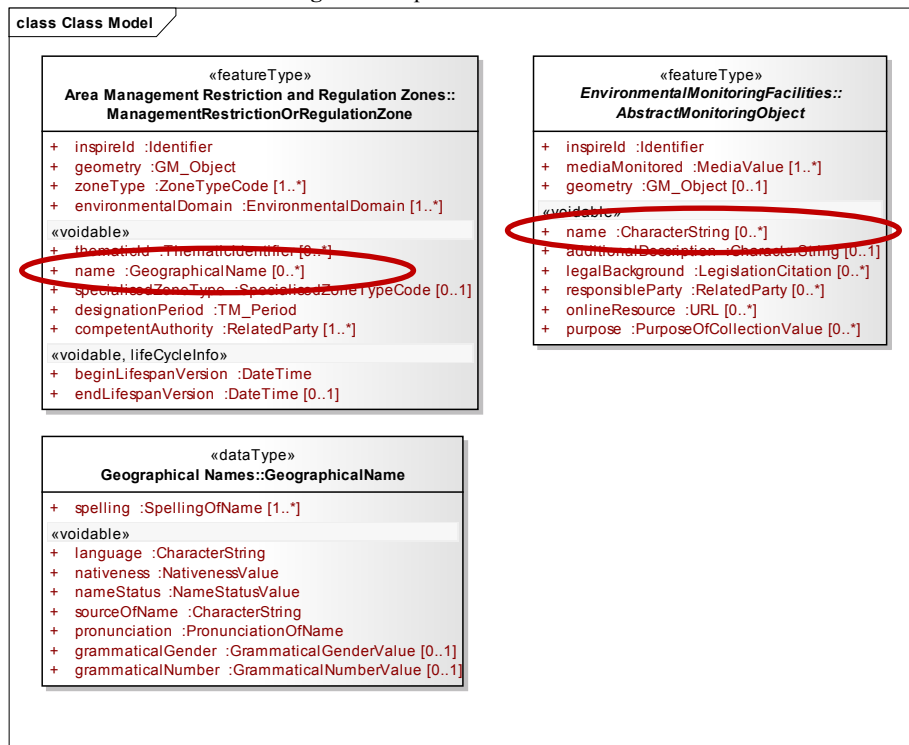
The rationale behind this difference is that within the TWG responsible for the Environmental Monitoring Facilities theme the decision was reached that the complexity of the Geographical name type was far greater than the requirements for Environmental Monitoring Facilities (see cost/benefit considerations in the INSPIRE data specification cycle shown above). However, this is sure to cause confusion in developers creating applications for multiple INSPIRE Themes.

4.2 INSPIRE Extensions - Loose Governance

Now that the INSPIRE data and service specifications have been finalized and are being implemented on both the national as well as the European levels, further extensions are being created of these base INSPIRE data models. An extension of the INSPIRE data models for European Air Quality e-Reporting has been finalized and is now operational[6]; the same approach is now being followed for various other environmental reporting obligations [6]. While these extended data models will be tailored to the requirements of a specific reporting obligation, they will undoubtedly be adding some similar concepts such as an European station code to the base INSPIRE models.

---

[6] http://www.eionet.europa.eu/aqportal

At present there is no mechanism in place to enable alignment of equivalent concepts across the various pending extensions of the existing data models.

In the example below we show the AQD_Station class developed for the European Air Quality e-Reporting and derived from the INSPIRE EnvironmentalMonitoring-Facility Class. Based on the requirements of the underlying air quality directive 2008/50/EC [7] and Commission Implementing Decision 2011/850/EU [8], in addition to the basic name provided by the INSPIRE EnvironmentalMonitoringFacility Class, a European Station Code must be provided. This attribute has been added to the AQD_Station definition as shown below.

Fig. 3. Example - INSPIRE Extension
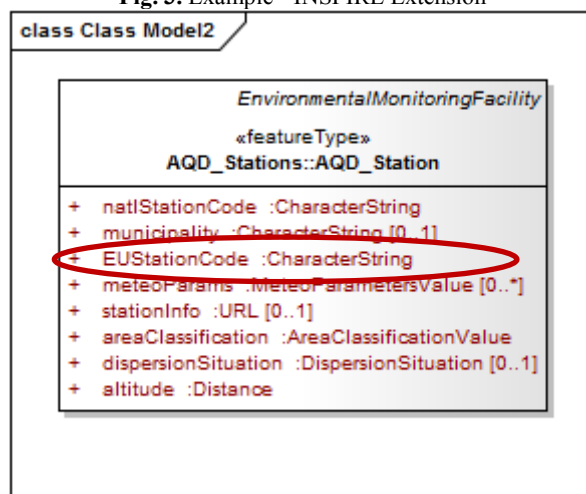
The process performed for European Air Quality e-Reporting shall now be repeated for various environmental reporting obligations, in each case the concepts ensuing from the legal requirements stemming from the relevant legislation shall be identified and the base INSPIRE data models will be extended accordingly. The likelihood that a concept such as a European Station Code will be required in other thematic domains is high; however, there is no mechanism in place to assure that this will be done in the same manner in each case, leading to subtle differences in these data specifications, and hindering reuse of mechanisms and code when working across thematic areas.

While this divergence could conceivably be managed by the European Environment Agency (EEA) when it pertains to the data specifications required in the environmental reporting domain, there is currently no mechanism in place for such coordination on extensions stemming from other areas or actors (i.e. national thematic extensions).

The INSPIRE Maintenance and Implementation Framework (MIF) is currently being set up between the European Commission and the Member States to guide further development of the INSPIRE process. An expert group called INSPIRE Maintenance and Implementation Group (MIG) with representatives of the INSPIRE national contact points has been established. Of the tasks identified under the Maintenance and

Implementation Work Programme (MIWP), the problems discussed here will be addressed by MIWP-14: Theme specific issues of data specifications & exchange of implementation experiences in thematic domains [9].

5 Lessons Learned

The examples above illustrate how despite best effort by all parties involved there is an inherent creep towards diversification in a data specification process. This can be seen as evidence that despite all the work to date on creating a process for the development of harmonized data specifications, there is still a necessity for further work in this area. In addition, this task becomes more challenging once the governance becomes looser as seen in the INSPIRE extension example.

Thus, further mechanisms must be put in place in order to avoid a new level of chaos to creep over the aligned core. Such mechanisms must be defined and put into place in the following areas:

- Governance: the INSPIRE data specification development is governed by the JRC, with increasing support from the EEA. However, this does not cover extensions ensuing from drivers not coming from the European administrative domain. Thus, mechanisms are required enabling a higher level of inclusion among a wider range of stakeholders.
- Processes: currently the INSPIRE maintenance process is being defined and various working groups are being created for this work. However, the same problem as pertains to governance is also active here, with little support for Non-European extensions. In addition to inclusion of a wider base of stakeholders, mechanisms will be necessary that facilitate the access of data stemming from outside parties. Also, as illustrated above, even with the rigid processes set up by the JRC for the base INSPIRE data specification work, discrepancies still occurred.
- Tools: Various web sites as well as a common repository for the UML data models provide support in finding agreement on necessary extensions to the INSPIRE data specifications. However, the alignment of concepts remains a manual process, depending on the meticulous perusal of the various specifications by humans. If tools were available that support the alignment between different concepts, this would make the alignment process much easier and less prone to human error.

Coming back from the more specific INSPIRE example detailed above to the general question of collaborative extension of environmental data models, it becomes clear that the challenges described in the INSPIRE process apply to other similar initiatives. Similar difficulties encountered in maintaining compatibility among extensions have been seen in the various other thematic data sharing initiatives mentioned earlier. The problems described become ever more difficult to handle the larger and more disparate the user community becomes. A great deal of effort has been put into the development of processes and supporting tools for collaborative data model development by various stakeholders in different thematic and geographic areas, but there is still a long way to go until this fully supports the requirements of the communities involved.

These must also be adapted to allow for thematic extension of existing standardized data models by diverse stakeholders without this leading to a new level of diversification.

Based on examples from the INSPIRE data specification process, we have seen how easily such discrepancies arise, both with tight governance and even more so with loose governance. Only if the problems identified are properly addressed, and the necessary steps taken will it be possible to provide a truly harmonized and usable data system across thematic domains both for now as well as for future requirements.

References

1. Katalin Tóth, Clemens Portele, Andreas Illert, Michael Lutz, Maria Nunes de Lima (2012) A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructures, JRC REFERENCE REPORT Retrieved from http://inspire.ec.europa.eu/documents/Data_Specifications/IES_Spatial_Data_Infrastructures_(online).pdf
2. 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) Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2007.108.01.0001.01.ENG . More available at: http://inspire.ec.europa.eu/
3. Craglia, M. (2010): Building INSPIRE: The Spatial Data Infrastructure for Europe. ARC News, 5-7. Redlands, California. Retrieved from http://www.esri.com/news/arcnews/spring10articles/building-inspire.html
4. JRC (2008) D2.6 Methodology for the development of data specifications Retrieved from http://inspire.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.6_v3.0.pdf
5. INSPIRE Generic Conceptual Model Retrieved from http://inspire.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4.pdf
6. Schleidt K. (2013) INSPIREd Air Quality Reporting - European Air Quality e-Reporting Based on INSPIRE. J. Hřebíček et al. (Eds.): ISESS 2013, IFIP AICT 413, pp. 439–450, 2013.
7. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Retrieved from http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32008L0050
8. COMMISSION IMPLEMENTING DECISION of 12 December 2011 laying down rules for Directives 2004/107/EC and 2008/50/EC of the European Parliament and of the Council as regards the reciprocal exchange of information and reporting on ambient air quality (2011/850/EU). Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:335:0086:0106:EN:PDF
9. MIWP-14: Theme specific issues of data specifications & exchange of implementation experiences in thematic domains Retrieved from : https://ies-svn.jrc.ec.europa.eu/issues/2136