

Shaping PortalU[®]/InGrid[®] to meet the INSPIRE requirements - First experiences and some considerations on how to efficiently organize the German spatial data infrastructure as INSPIRE node

Dr. Martin Klenke, Dr. Fred Kruse, Stefanie Uhrich, Christiane Giffei, Sybille Peters¹

Abstract

Actual efforts to fit the ISO 19115 / ISO 19119 compliant data model of the German *Catalog of Data Sources* to the INSPIRE requirements formulated by the Draft Implementing Rules for Metadata (Version 4) are presented. Moreover, some considerations on how to set up an efficient German INSPIRE metadata infrastructure are discussed.

1. Introduction: UDK, PortalU[®]/InGrid[®] and INSPIRE

In Germany, the *Catalog of Data Sources* (Umweltdatenkatalog, UDK) is in use in most federal and state environmental agencies in order to describe and manage environmental data and information. Over the last decade, the UDK data model has evolved to a quasi-standard for referencing environmental information and data. The UDK metadata subset describing geospatial data and services is modelled following the international standards ISO 19115 and ISO 19119. The *Catalog of Data Sources* is fully integrated into the German Environmental Information Portal, PortalU^{®2} and its underlying software InGrid[®] (Klenke et al. 2006).

In the context of the INSPIRE³ directive, the Implementing Rules (IRs) for Metadata are of particular significance for the European geospatial metadata communities. As a member of the *INSPIRE Drafting Team Metadata*, the Coordination Center PortalU has been involved in the formulation of the Implementing Rules which have been approved by the INSPIRE Committee on May 14th, 2008. As the first set of a series of INSPIRE Implementing Rules, the Metadata IRs are supposed to be finally published in the second half of 2008 and will then directly become effective as applicable law in the member states. In the second part of our paper, we give a brief overview of UDK data model extensions we found necessary to satisfy INSPIRE demands. Probably, other ISO 19115/19119 compliant data catalogs will face roughly the same tasks.

With respect to the INSPIRE Discovery Services, the commitment of the Drafting Team *Network Services* to the OpenGIS[®] Catalogue Services Specification 2.0.2 - ISO Metadata Application

¹ Coordination Center PortalU at the Lower Saxony Ministry of Environment and Climate Protection

² <http://www.portalu.de>

³ <http://www.ec-gis.org/inspire>

Profile Version 1.0 has implications on performance and topology of the involved catalog services. Hence, currently there are intense discussions going on within the German geospatial community under the roof of GDI-DE⁴ regarding the question on how to establish a suitable German metadata node and service topology for INSPIRE. In the third part of our paper, we present some considerations on this topic based on discussions and experiences within the GDI-DE pilot project *Geodatenkatalog-DE*.

2. INSPIRE-induced UDK Data Model adjustments

The following sub-chapters describe the modifications and extensions applied to the ISO-compliant UDK data model in order fulfil INSPIRE demands derived from the *INSPIRE Draft Implementing Rules for Metadata (Version 4)*.

Conformity (Draft: 2.2.7) New mandatory tripartite 1:n field group consisting of *Specification title*, *reference date* and *Degree of conformity*. Content shall provide at least information on the degree of conformity with the Implementing Rules provided in Art. 7-1 of the INSPIRE Directive.

Unique resource identifier (Draft: 2.2.1.5) New mandatory identifier for the data source described by the metadata object. Concrete implementation is not yet clear and will be provided later by the *Data Specifications* IRs. At first, we will use a catalog identifier and the UUID of the metadata object separated by a colon, e.g. UDK-NI : <UUID>.

Constraint related to access and use (Draft: 2.2.8) New, two-column table with the headers *Limitations on public access* (fixed code list) and *Conditions applying to access and use*. Per line, to complete both fields becomes mandatory. Existing UDK/ISO-Fields can be re-mapped to provide the information.

Lineage (Draft: 2.2.6.1) Information on process history and quality of the spatial dataset. The equivalent optional ISO 19115-entity (LI_Lineage) becomes mandatory for INSPIRE-conformal catalogs.

Spatial data service type (Draft: 2.2.2.2, Annex B.3) This information is mandatory for spatial services. It has to be filled by an underlying fixed code list which is more restrictive than the mandatory ISO 19119 service type attribute. For the UDK, this information can be automatically extracted.

Spatial resolution (Draft: 2.2.6.2) In the UDK, this attribute is optional for spatial data sets and data set series. In addition, it will be introduced to Spatial Data Services.

⁴ <http://www.gdi-de.org/>

Metadata point of contact (Draft: 2.2.10.1) Here, the most important difference between ISO 19115 and INSPIRE is the switch from optional to mandatory for the attribute *E-Mail-Address*. In order to apply for this INSPIRE-demand, it is necessary to identify an email-address for all metadata objects which do not have already at least one. As the UDK address metadata is structured hierarchally and at least an email-address must exist for the catalog administrator, this can initially be done automatically for existing UDK metadata objects.

Keyword, Spatial Data Service (Draft: 2.2.3 Keyword, Annex B.4) In ISO 19119, the keyword-attribute is optional. For INSPIRE Spatial Data Services, one entry from the extensive code list shown in Annex B.4 of the IRs Metadata gets mandatory. For existing UDK metadata, this information will initially be automatically generated from the Spatial Data Service Type.

Keyword (Draft: 2.4) ISO 19115 requires at least one keyword per resource. There is no need to use formalised words. For INSPIRE, at least one keyword from the General Environmental Multi-Lingual Thesaurus (GEMET⁵) becomes mandatory. At present, for UDK data at least three keywords from the UMTHESES⁶-Thesaurus are mandatory. For existing UDK objects, GEMET keywords will automatically identified from the assigned UMTHESES keywords by means of the SNS⁷ webservice. In the future, an SNS-based thesaurus assistant will support UDK metadata editors in order to find appropriate GEMET keywords for their resources.

3. The German Spatial (Meta-)Data Infrastructure and INSPIRE

From the current state of discussion in the Drafting Team Network Services one can expect, that the *OpenGIS® Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile (1.0.0)*⁸ becomes the basis of INSPIRE metadata flows. For catalog servers, high performance requirements are formulated in the *Draft Implementing Rules for Discovery and View Services (IRI) Version 2.0*. For instance, the response time for 250 discovery service records should be less than 3 seconds in 90% of the servers uptime. Moreover, the minimum number of simultaneous discovery service requests served according to the performance requirement shall be 100.

The GDI-DE pilot project *Geodatenkatalog-DE* aims to develop a prototype of the central German metadata node. Main challenge is the development of an efficient topology of the relevant German catalog services actually being extremely heterogeneous in terms of content structure (theme-based, administrative, etc.) and technical implementations.

⁵ <http://www.eionet.europa.eu/gemet>

⁶ <http://www.umweltbundesamt.de/uba-info/dokufabib/thes.htm>

⁷ <http://www.semantic-network.de/>

⁸ http://portal.opengeospatial.org/files/?artifact_id=21460

With particular respect to the INSPIRE performance requirements, harvesting and indexing processes will be necessary to consolidate the content of distributed, partly cascaded German catalogs for INSPIRE. For information brokers like the German GeoPortal.Bund⁹ and PortalU[®], suitable functionalities have to be realised. For the PortalU[®]-software InGrid[®] we plan to implement the following functionalities: Harvesting of catalog contents by means of a wildcard-*GetRecords*-CSW-request (*brief* result set) followed by a *GetRecordByID*-Request with *full* result set. Responses will be saved in a server-side XML-File ready for indexing using an index field to XPath Mapping. Therewith, responses to discovery service requests from the INSPIRE geoportal can be answered rapidly from the index with no need to delaying requests to the original data sources.

Particularly in multiply cascaded catalog topologies duplicates can be generated during harvesting. Those have to be filtered by the broker using the unique resource IDs and timestamps.

An open question is how to pool the CSW AP ISO query capabilities with the demands stated in the INSPIRE Directive Article 11(2). For example, INSPIRE demands *conformity* (see above) as queryable although ISO/CSW AP has no exact equivalent. However, to completely map the INSPIRE metadata model to ISO 19115/19119 should be attempted. That would avoid the need for an extended “INSPIRE-Profile” which probably generates new problems for catalog implementations and overall catalog interoperability.

4. Summary

While the INSPIRE IRs are taking more shape there are still huge uncertainties regarding their practical implementation. Basically, the existent INSPIRE-relevant German catalogs seem to be well-prepared as they usually comply to the ISO 19115/19119 and commit themselves to CSW AP ISO 1.0 interfaces in accordance with the GDI-DE *Architekturkonzept*.¹⁰

The topology set-up of German catalog services for INSPIRE, however, still remains open up to now. Experiences with the implementation of our existing catalog interfaces point out that, to balance the benefits and problems of the different conceivable topology approaches and to find a profound solution heavily depends on prototyping and testing under real world conditions. Hence, application-oriented projects, like Geodatenkatalog-DE, are significant at present.

5. Literature

Klenke, Martin; Kruse, Fred; Lehmann, Hanno; Riegel, Thomas; Vögele, Thomas: InGrid[®] 1.0 – The Nuts and Bolts of PortalU[®]. EnviroInfo 2006, Graz, 2006.

⁹ <http://geoportal.bkg.bund.de>

¹⁰ http://www.imagi.de/de/download/Managment_A_Konzept.pdf