

Metadata Creation and Management of Distributed Data Catalogs with PortalU[®] and InGrid[®] 1.1

Thomas Vögele¹, Martin Klenke¹, Fred Kruse¹

Abstract

Interoperable online catalogs of geospatial data and services will play an important role in the implementation of the EU-INSPIRE Directive. This includes catalogs of environmental geospatial resources. With the Umweltdatenkatalog (UDK) and the German Environmental Information Portal (PortalU[®]), the environmental administration in Germany already disposes of a technical infrastructure in support of INSPIRE process. This infrastructure is continuously adapted to international standards and INSPIRE implementing rules to ensure interoperability with national and European spatial data infrastructures with respect to metadata schemata as well as to catalog interfaces. The infrastructure will be complemented by a new online tools for metadata creation and quality control. This tool will reduce the effort required for the management of up-to-date, comprehensive and interoperable data catalogs.

1. Introduction

On May 15th 2007, EU-Directive 2007/2/EC on an *Infrastructure for Spatial Information in Europe* (INSPIRE) entered in force (EU 2007). The intent of the INSPIRE-Directive is to establish a European Spatial Data Infrastructure (ESDI) that enhances the exchange and the use of geospatial data throughout Europe. The scope of INSPIRE is not limited to geospatial reference data like topographic maps, areal photos, catastrophe information, and the services needed to process these data: Because the thematic focus of INSPIRE is the environment, “geospatial data” addressed by the Directive do also include information on protected areas, soil quality and other environmental themes.

Due to this strong thematic focus on the environment, the INSPIRE-Directive will have far-reaching implications for the environmental administrations in the EU member states: Environmental geospatial data are typically produced by public agencies on all levels of the administrative hierarchy. They also host most of the services needed to visualize and further process these data. In federally-organized member states like Germany, a large section of the actual data producers and service providers can be found on the lower level of the hierarchy, i.e. in the federal states and also the municipalities. Many of them will have to be included in the INSPIRE process and in the envisioned European Spatial Data Infrastructure. Given the large number of organizations and the inherent heterogeneities in the administrative structures of the EU member states, this goal will not be an easy one to achieve.

A central part of the new European Spatial Data Infrastructure, and one that will have to be developed by the European Member States during the first phase of the technical implementation of INSPIRE, are comprehensive and interoperable data catalogs. Ultimately, these catalogs will have to provide metadata for all geospatial data and services that are addressed by INSPIRE and that fall under the thematic categories listed in the INSPIRE annexes. Within the ESDI, these catalogs will have to be accessible through online portals and provide information for the discovery, evaluation and use of the referenced resources.

2. An Infrastructure for Online Metadata Research and Retrieval

In Germany, the environmental administration is well prepared for this first phase of the INSPIRE challenge: With the *Catalogue of Data Sources* (Umweltdatenkatalog - UDK), a comprehensive electronic

¹ Coordination-Center PortalU, Lower-Saxony Ministry for Environment, Archivstrasse 2, 30169 Hannover, Germany

data catalog has been functional for more than 10 years (Swoboda, 2000). Since June 2006, the UDK is fully integrated into PortalU[®] (<http://www.portalu.de>), which is the new central online portal for publicly-held environmental information, data and services in Germany (Vögele et al., 2006). Through PortalU[®], UDK-metadata can be researched and visualized with the help of a user-friendly web-interface and powerful search tools. These range from a systematic “catalog of environmental themes” and a “google-type” simple fulltext-search facility to a complex, thesaurus- and map-supported spatio-temporal search tool.

Currently the UDK (i.e., the PortalU[®] metadata component) is used mainly by federal and state environmental agencies to manage environmental data and information. In Germany, the UDK data-model has evolved into a quasi-standard for the referencing of environmental information and data. Because a large percentage of environmental data do have a spatial component, the UDK does not only describe documents, publications, projects and other types of environmental information objects, but also environmental geospatial data and services: Among the approximately 12.000 data records in the UDK (as of May 2007), more than 2.500 point to digital maps and/or areal photos. In addition, a small but growing section of the UDK-records relate to services. Currently, almost 650 services are referenced².

The UDK data-model had been harmonized with the currently most relevant international standards for geospatial metadata and services, i.e. ISO 19115 and ISO 19119. This enables the UDK to reference environmental geospatial data and services in a standardized, harmonized and interoperable way. Because the long-term operation, maintenance and development of PortalU[®] and the UDK are put on a sound financial and organizational basis by an administrative agreement between the participating state- and federal agencies, it is possible to continuously improve and update the system and to keep up with the evolution of the (ISO) standards as well as with upcoming specifications, like the INSPIRE Implementing Rules for Metadata. This makes it possible to ensure that the UDK metadata-model and the content of the UDK/ PortalU[®] data catalog remain interoperable with the content of other geodata catalogs within the national German Geodata Infrastructure (GDI-DE) and the envisioned European Spatial Data Infrastructure (ESDI).

In order to allow the content of the UDK data catalog to be exchanged within a geodata infrastructure via interoperable catalog interfaces, PortalU[®] was equipped with an interface that adopts a profile of the OGC Catalog-Service-Web (CSW) specification (OGC 2005, OGC 2007). Through this interface, PortalU[®] and the UDK will become searchable by Geoportal.bund (<http://geoportal.bkg.bund.de>), the central information portal of GDI-DE. In the future, the interface may also serve to connect directly to geoportals within the ESDI.

3. A New Tool for Online Metadata Management

Highly accessible and user-friendly tools to retrieve metadata for the discovery, evaluation and use of geospatial data and services are important cornerstones of a successful data catalog application. However, they are of little use if the data catalog is not kept up-to-date, or if metadata records are missing. Therefore another important prerequisite for a comprehensive data catalog is the possibility to create, update and manage metadata as effectively as possible. Analogous to web-based applications for research and retrieval, online tools for data entry, quality-control and catalog management are needed. The UDK, as well as the current metadata module of PortalU[®], still rely on a desktop data-entry application (UDK 5.0). This makes the creation of metadata and the management of the (distributed) UDK catalogs unnecessarily complex and cumbersome.

To overcome this drawback, a new tool for online metadata entry and management is currently under development. The next version of the PortalU[®]-software InGrid[®] will contain a module that enables data providers to handle the entire workflow of metadata acquisition, metadata quality control, and metadata

² Not all of these services are geo-services in the sense of an Web-Mapping- or Web-Feature-Service as specified by the OGC.

research through a user-friendly Internet application (Klenke, 2006). Technically, this application is based on AJAX technology (Garrett, 2005), which means that it is fully browser-based and does not rely on the download of plug-ins and other software.

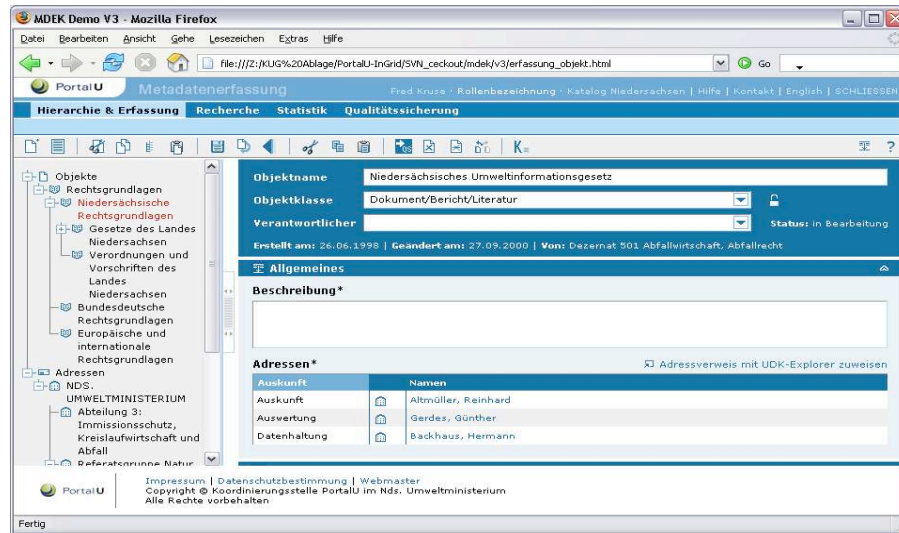


Figure 1: UDK/InGrid Online Data Entry Tool (design study)

In addition to a user-friendly web-interface (Figure 1), the new online data-entry tool (InGrid® 1.1) will feature a number of technical improvements:

- *Automatic consistency checks:* Before a new metadata record can be entered in the workflow, the system checks for the consistency and availability of a set of core-metadata. This set represents the minimum of information that has to be provided as specified by the ISO 19115 /19119 core data set and the INSPIRE metadata for discovery as specified in the INSPIRE Implementing Rules for Metadata.
- *Configurable data-entry forms:* The administrator of a data catalog may specify which metadata elements (in addition to the core elements) have to be filled out. Only those fields that are relevant are shown to the user. This is an important feature to implement provider-specific metadata profiles.
- *Navigation by hierarchy:* Navigation through and visualization of the (hierarchical) structure of the UDK metadata model is supported through an interactive tree view.
- *Thesaurus-aided keyword annotation:* The selection of keywords and place names for thematic and spatial annotation of metadata records is supported by SNS (Bandholtz, 2003), an environmental and spatial thesaurus provided as an online service by the German Umweltbundesamt (UBA).
- *Automatic annotation of keywords:* A keyword-assistant can be used to semantically analyze metadata records (i.e., title, abstract and other descriptive fields) and/or online documents. The assistant proposes matching keywords on the basis of the SNS thesaurus mentioned above.

- *Manual quality control*: Before it is stored in the data catalog, each new data record may be submitted to an (optional) quality control workflow. This workflow supports the peer-review of a metadata record by a dedicated quality control person.
- *Automatic actuality control*: The quality-control module is also used to ensure the actuality of individual data records. Based on the initial time-stamp of a record, quality control actions are automatically triggered at specified time intervals.
- *Multi-session processing*: The metadata creator may work on a metadata record in a number of subsequent sessions. After each session, a metadata record may be saved temporarily and re-loaded at the beginning of the next session.
- *Rights-based administration*: The administration of multiple data creators that simultaneously contribute to one single data catalog is handled by a straight-forward rights-based scheme. Each metadata creator is given writing-access only to his or her sector of the data catalog.

4. Summary and Conclusions

The pending implementation of the EU-INSPIRE Directive will force public administrations throughout Europe to invest into the development of online metadata catalogs for geospatial reference data and services, as well as for environmental geospatial resources. With the Umweltdatenkatalog UDK and PortalU[®], the environmental administration in Germany has already set up an electronic data catalog and an online portal to reference and research, among others, environmental geospatial data and services. Both tools are in-line with international standards and new implementing rules provided by INSPIRE. Standardized catalog interfaces will ensure interoperability with the German and the European spatial data infrastructure.

To be useful for the discovery, evaluation and use of geospatial resources, a data catalog has to be up-to-date and comprehensive. With InGrid[®] 1.1, a new tool to improve the content of the UDK/PortalU[®] data catalog will be available soon. This new online metadata entry tool will simplify and streamline the processes of metadata acquisition, quality control, and the long-term maintenance of UDK/ PortalU[®] data catalogs. It is expected that this will reduce the cost of metadata creation and maintenance, and lead to comprehensive, up-to-date and interoperable catalogs of environmental geospatial data and services.

Bibliography

- Bandholtz, T. (2003): Erstellung eines semantischen Netzwerkservice (SNS) für das Umweltinformation-
netz Deutschland (gein[®]). Abschlussbericht, Umweltbundesamt.
- EU (2007): Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 estab-
lishing an Infrastructure for Spatial Information in the European Community (INSPIRE), Official
Journal of the European Union, Volume 50, 25 April 2007, L 108, ISSN 1725-2555.
- Garrett, J.J. (2005): Ajax: A New Approach to Web Applications. Adaptive Path LLC, 18. Februar 2005
- Klenke, M., Kruse, F., Lehmann, H., Riegel, T., Vögele, T. (2006): InGrid 1.0 – The Nuts and Bolts of
PortalU, Proceedings of EnviroInfo held in September 2006, Graz, Austria.
- OGC (2007): OpenGIS[®] Catalogue Service Implementation Specification Version 2.0.2.
Web: http://portal.opengeospatial.org/files/?artifact_id=20555 (25.04.2007).
- OGC (2005): ISO19115 / ISO19119 Application Profile for CSW 2.0. Version 0.9.2
Web: http://portal.opengeospatial.org/files/?artifact_id=6495 (25.04.2007).
- Swoboda, W. et al. (2000): Harmonisierter Zugang zu Umweltinformationen für Öffentlichkeit, Politik
und Planung: Der Umweltdatenkatalog UDK im Einsatz. In: K. Tochtermann, W.-F. Riekert

(Eds), Proceedings of the 14th International Symposium „Computer Science for Environmental Protection”, Bonn.

Vögele, T., Klenke, M., Kruse, F., Lehmann H., Riegel, T. (2006): Easy Access to Environmental Information with PortalU, Proceedings of EnviroInfo held in September 2006, Graz, Austria.