The UDK and ISO 19115 Standard

Oliver Karschnick[1], Fred Kruse[1], Stephani Töpker[1], Thomas Riegel[1], Marco Eichler[2], and Sven Behrens[3]

Abstract

The German Environmental Data Catalogue UDK (German: Umweltdatenkatalog) is the standard tool for public access to environmental metadata in Germany and Austria. Considering the necessity of presenting more data from different sources in a homogeneous way, the UDK needs a normative data model for exchanging easily data with other databases like GeoMIS.Bund. In order to achieve this aim the UDK data model is extended to be compatible with the ISO 19115 standard (International Organization for Standardization) for geographic Information, finally aiming to fulfill the Directive of the European Parliament and of the Council on public access to environmental information. In this paper an actual review about the UDK is given and the development of the ISO 19115 compatibility is presented.

Zusammenfassung

Der Deutsche Umweltdatenkatalog UDK ist die Standardanwendung für den öffentlichen Zugang zu Umweltmetadaten in Deutschland und Österreich. Aufgrund der Notwendigkeit, immer mehr Umweltdaten aus unterschiedlichen Quellen einheitlich darzustellen, benötigt der UDK ein normiertes Datenmodel, so dass der Datenaustausch mit anderen Systemen, wie z.B. GeoMIS.BUND ermöglicht wird. Um dieses Ziel zu erreichen, wird das UDK Datenmodell gemäß des ISO 19115 Standards (International Organization for Standardization) erweitert. In diesem Artikel wird ein aktueller Überblick über den UDK gegeben und die Anpassung an die IS0 19115 vorgestellt.

[1] Niedersächsisches Umweltministerium, Koordinierungsstelle UDK, Archivstr. 2, D-30169 Hannover, kug@numis.niedersachsen.de, -http://www.udk-gein.de
[2] BVG mbH, Menkestr. 7, D-26419 Schortens, udk@bvg-fri.de, http://www.bvg-fri.de
[3] disy Informationssysteme GmbH, Stephanienstr. 30, D-76133 Karlsruhe, behrens@disy.net, http://www.disy.net

1. Introduction

The Environmental Data Catalogue (German: Umweltdatenkatalog UDK), introduced in 1991, is the standard tool in Germany and Austria to allow an efficient and easy public access to environmental data (Swoboda/Kruse 1999, Swoboda et al. 2000). Different programs have been developed for the collection and administration of metadata and its presentation through the Internet. Using custom-made interfaces based on XML (Kruse et al. 2001) and the SOAP-Protocol it is possible to search other metadata systems and merge these results with those of the UDK in a standardized way. Furthermore, the UDK can be used for search and data-retrieval by other information systems like the German Environmental Information Network (Umweltinformationsnetz Deutschland gein®). Vice versa, a metadata component of gein® is the UDK.

In order to achieve the possibility of normative data exchange with other information systems, like the German Broker for geospatial metadata GeoMIS.Bund, the UDK is extended to be compatible with the ISO 19115 (International Organization for Standardization) standard (International Organization for Standardization 2003). To achieve this aim, the UDK data model will be enhanced according to the requirements of that ISO norm.

2. The UDK

The UDK has been developed to collect, administrate and retrieve environmental metadata. This data should answer the question: *Who* holds *where what* kind of data in *which* format?

Different tools have been developed for the various tasks: for the collection of metadata a Windows based UDK (Win-UDK) is available. In addition this software can be used to search and administrate the database, too. Furthermore, a web-based WWW-UDK was developed for public retrieval of data via the Internet or intranet. Actually, nearly 30,000 datasets (and in addition objects from the Bavarian Catalogue, which is connected) can be searched with the Web component of the UDK in Germany. In Austria almost 15,000 datasets have been collected so far.

2.1 Windows based collection and retrieval of metadata

The Windows UDK, current Version 4.3, is used to collect and retrieve environmental metadata. Usually the experts responsible for the data themselves realize the data input. In this way, a high quality of data can be achieved. Based on the decentralized input a decentralized software as is necessary as well.

The Win-UDK 4.3 is available in a standalone ('E-Version') or a client server version ('CS-Version'), for both of which a three-layer architecture is used. For the communication between the client and the server through the local area network (LAN) or the Internet the SOAP technique (Simple Access Object Protocol) is used (Kruse/Eichler 2001). A special update service guarantees that all databases obtain the latest data. In addition an update service for the client software is integrated to provide up to date identical software.

Both, the E and CS version are developed with Visual Basic 6 running on actual Windows operating systems. The databases Oracle, MS-SQL-Server, Access, and Informix are supported. In addition, the Win-UDK contains a XML-based interface to provide an import of data from other information systems.

2.2 Web-based retrieval of environmental metadata

The Web-UDK allows every Internet user to retrieve the data which was collected by the use of the Win-UDK. Actually, the Virtual UDK (V-UDK, can be accessed on http://www.umweltdatenkatalog.de) allows retrieval accessibility to the databases of the UDK-data collected in Germany. At the moment there are about 25,000 objects and 5,000 addresses included, a number which is rapidly growing.

The Web-UDK is a multi layer client server application developed in Java and provides two user interfaces: A Servlet/JSP based web application can be used with a standard web browser. The UDK Server manages client requests, queries the underlying database and maps the result to a UDK business object, which is then printed for the user. The Virtual UDK has been realized by adding an extra layer for distributing the queries and collecting the results. The V-UDK Server is communicating with the distributed UDK Servers by the means of Java RMI and the Bavarian Object Data Catalogue (UOK) utilizing SOAP technology. Being the metadata component of the German Environmental Information Network *gein*®, the V-UDK is connected to *gein*® by using the GEIN-2000 XML-Profile.

2.3 Data Model

The metadata records are stored as so-called UDK objects, the attributes of which are e.g. object identification number, object name, general descriptions, key words as well as time and spatial coverage.

The objects are assigned to 6 different UDK object classes. These classes are labeled as:

1. dataset / database
2. service / application / information system
3. document / report / literature
4. geographical information / map
5. organizational unit / task
6. project / programmes

Addresses of the administrations and objects being assigned to one catalogue are distinguished. Based on the identification number objects and addresses can be linked hierarchically (parent-child relation) and in parallel (object-object relation), which is a special feature of this data model. Furthermore, it is possible of course to include several catalogues into one database.

3. The UDK and the ISO 19115 Standard

Operating in a steadily growing heap of environmental data, it is inevitable to ask for an increasing functionality of the UDK in terms of communicating with other environmental information systems, e.g. geographic information systems like ArcGIS or GeoMIS.Bund. A communication is only possible with a standardized structure of data. A first step aiming for such a data structure was to implement standards like XML and SOAP. The next task is to gain the compatibility to the ISO 19115 standard.

The ISO 19115 standard is a schema, which standardise the description of geographic information and services like the identification, the extent, the quality, the spatial and temporal schema, the spatial reference, and the distribution of digital geographic data. In this standard mandatory, conditional and optional metadata sections are specified as well as metadata entities and elements, a minimal set of which is defined for the most important information needed. This data are called ‘core data’. Of course there is the possibility to extend this data model to fit special needs.

Only a part of the UDK data, compared to the ISO data, is geographical data, and only a part of it is included in the ISO core data. These relations are plotted in Figure 1. In the left plot of this figure the ISO 19115 model is shown, with a possible extension. In the right plot all components of the UDK data are shown (qualitative) in addition to the ISO 19115 model. Only the data of the object class “geographical information/map” is part of the ISO 19115, and only a fraction of it is part of the core data.

Taking these points into consideration the ISO 19115 compatibility of the UDK is being achieved by following four steps:

1. Comparing all data fields of the UDK object class "geographical information/map" to the ISO 19115 components.

2. Expand the UDK data model concerning the core data: There are few fields in the core data which are not covered by the old UDK data model.
3. Expand the UDK data fields beyond the core data: The expansion of the UDK data model is not only limited to the core data. To fulfill the user requirements it is necessary to expand the UDK data model beyond the core data of the ISO.
4. Convert obligations/conditions: Some components of the UDK data model are part of the ISO standard, however, their obligation or condition ("mandatory", "conditional", "optional") is different. This has to be adapted.

There is no need to expand the ISO model, because there are no fields in the UDK model, which are not covered already by ISO components.

The expansion of the UDK data model can easily be described by comparing the left and right plot of Figure 1. The differences are shown in more detail in Figure 2, where only the geographical information, i.e. without the other classes of the UDK data model, is plotted. The components of the ISO 19115 and the core data is shown the same way as in the other figures. The black area shows the expansion of the UDK data model to reach the ISO conformity. In addition to the expansion, so that the core data is implemented, other components have to be added, so that all

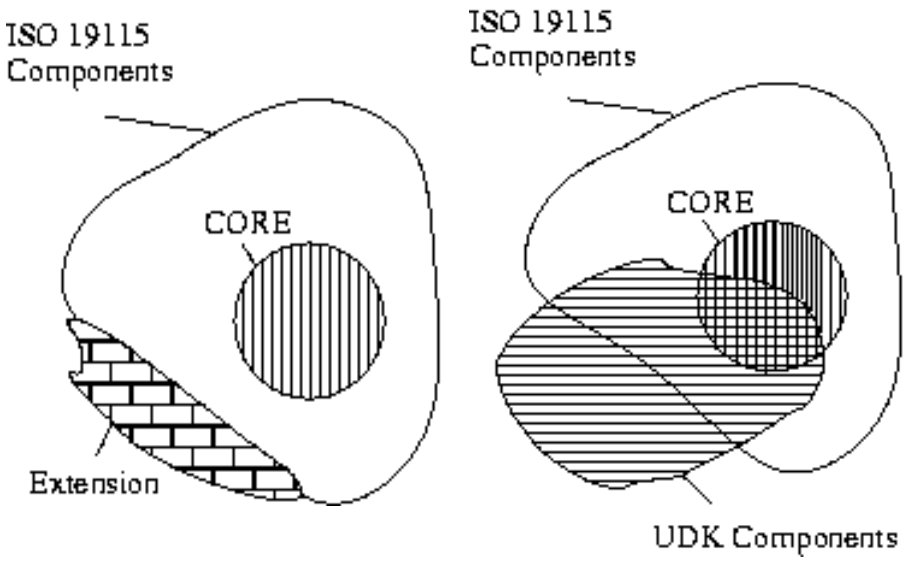

Figure 1: The ISO 19115 data components plotted qualitatively (left): The most important data is located in the core data. An extension to adopt the components on special needs according to special rules is possible. : The (qualitative) picture of the components of the UDK data as it is before the data model adoption compared to that of the ISO 19115 (right): Only the data of the object class "geographical information/map" is part of the ISO 19115, and only a fraction of it is part of the core data.

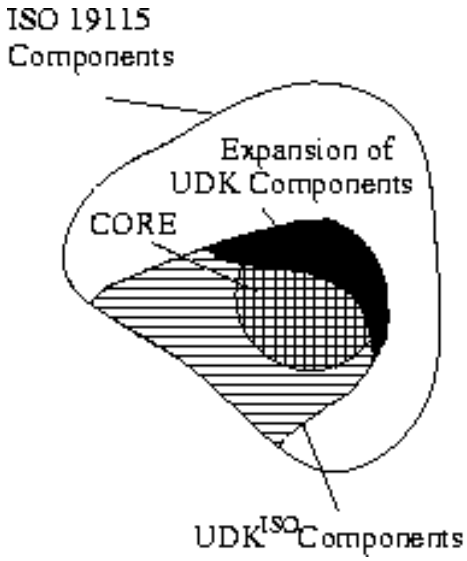

Figure 2: The (qualitative) picture of the UDK^{ISO} object class "geographical information/map" as it should be after the data adoption compared to that of the ISO 19115. The black area shows the components about which the UDK components have to be expanded to reach ISO conformity.

user requirements are fulfilled.

The model of the UDK data before the adoption compared quantitatively to that of the ISO 19115 is leading to the following result: The UDK data (81 fields over all) which is located outside of the ISO boundary (22 fields) belongs to object classes not assigned to geographical information. Whereas the fields of the UDK object class "geographical information/map" (59 fields) is in principle already part of the ISO 19115 components. However, only a fraction of this data is part of the core data (14 fields, i.e. fraction of 64 % of the core data), which is needed to be added at least to reach ISO conformity for the UDK.

4. Conclusion

The UDK is the standard metainformation system and management tool for environmental metadata in Germany and Austria. The UDK communicates already successfully with the Bavarian data catalogue UOK which can be searched by the Vir-

tual UDK. In addition, the German Environmental Information Network *gein*® (http://www.gein.de) searches the Virtual UDK.

Considering the great increase of interest and important environmental data throughout the internet, the necessity to fulfill the Directive of the European Parliament and the Council on public access to environmental information the need to communicate with more and different information systems is growing. Especially the management of geographical metadata will play a more important role in the future. The requirement to realize a communication amongst different systems is a standardized data model. Such a model is defined by the ISO 19115 international standard for geographical metadata, to which the UDK is being adopted and thus will be a part of a globalized environmental information network.

References

International Organization for Standardization (2003). ISO 19115:2003 Geographic Information Metadata

Kruse, F., Eichler, M.,et. al. (2001a): Die XML-Schnittstelle des UDK 4.2 -Der Schlüssel zur Integration von Umweltanwendungen, Tagungsband des 4. Workshop des GI-Arbeitskreises Hypermedia im Umweltschutz und 3. Workshops der GI-Initiative Environmental Markup Language, Ulm, Germany

Kruse, F., Eichler, M. (2001b): Der Windows UDK 4.3 auf Basis der SOAP Technologie, in: M. Hilty, P.-E. Gilgen (Eds). Proceedings of the 15th International Symposium Informatics for Environmental Protection „Sustainability in the Information Society", Zurich, Switzerland

Kruse, F., Karschnick, O., et al. (2002): The UDK - Present Status and Future Development An Overview, in: K. Tochtermann, W. Pillmann (Eds), Proceedings of the 16th International Conference "Informatics for Environmental Protection". Vienna, Austria

Swoboda, W., Kruse, W., et al. (1999): The UDK Approach: the 4th Generation of an Environmental Data Catalogue Introduced in Austria and Germany, Proceedings of the 3rd IEEE Meta-Data Conference, Bethesda, Maryland, USA

Swoboda, W., Kruse, 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, Germany