EnviroInfo 2005 (Brno)
Informatics for Environmental Protection - Networking Environmental Information
 ISBN: 80-210-3780-6

Environmental Data Catalogue: Implementation and Data Quality Issues

Peter Fischer-Stabel[1], Rolf Krieger[1], Kerstin Schäfer[1] and Stefan Wannemacher[1]

Abstract

In Germany the environmental data catalogue was designed as a meta information system for retrieving environmental information held by public authorities. It is providing its user with information about who is having which information at what location. In this paper we discuss some issues concerning the implementation of the environmental data catalogue at public authorities and the management of data quality.

1. Introduction

The environmental data catalogue (UDK: **U**mwelt**d**aten**k**atalog) is a software system designed for managing meta data about environmental information held by German public authorities. It plays an important role in the implementation process of the Directive 2003/4/EC on public access to environmental information. Examples of environmental information objects include databases, documents, information systems, maps, projects, etc. on the state of the environment. The meta data record describing an environmental information object is called UDK object (Swoboda 1999). Essential in implementing and maintaining a meta information system is the data collection and entry process and the data updating process. Both have a strong influence on the meta data quality. Thereby, we have to face the fact that there is a huge number of data holders located at different distributed organisational units. Therefore, meta data collection and the periodical updating of meta data records are very complex tasks. In this paper we describe processes and tools supporting both tasks. Our results are based on an empirical study about the implementation of the environmental data catalogue at public authorities of the German states of Rheinland-Pfalz and Saarland.

2. Implementation

The implementation of the environmental data catalogue is a complex process comprising a lot of activities. Important activities are:

1. Definition of a taxonomy for classifying environmental information (structure tree).
2. Meta data collection concerning environmental information held by the considered public authorities, the definition of the corresponding UDK objects (meta data records) and the entry of the UDK objects.
3. Insertion of the UDK objects into the structure tree.
4. Definition of a process for updating the meta data periodically.

One important objective of the data collection is that meta data records are available for all environmental information objects published on the Web site of a public authority. Moreover, each meta data record must

[1] University of Applied Sciences Trier, Umwelt-Campus Birkenfeld, P.O. Box 1380, D-55761 Birkenfeld, Germany

refer to the corresponding environmental information object provided on a Web page as described in Fig. 1. The UDK offers an attribute to define this relation between environmental information object and the corresponding UDK object.

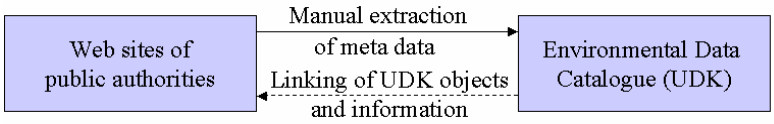

Fig. 1: Relation between environmental information and UDK objects.

In the following subsections we analyze the activities listed above in more detail.

2.1 Definition of the Structure Tree

Important for managing and searching UDK objects in an efficient way is the definition of a structure tree. The structure tree corresponds to a taxonomy for environmental information. Based on the taxonomy UDK objects have to be classified and inserted in the structure tree. There are different approaches for classifying UDK objects, e.g.

- **Subject-oriented Classification**
  UDK objects are classified according to environmental subjects like climate, air, soil, water, waste, etc. Each node of the structure tree corresponds to a subject. The subjects are arranged in hierarchical way. In (Koordinierungsstelle UDK 2002) a taxonomy is proposed that is usually customized according to the requirements of a public authority. For instance in some cases additional subjects are needed.

- **Organisation-oriented Classification**
  UDK objects are classified according to the organisational structure of the public authority considered. For instance, the root node of the structure tree is the Ministry. The departments of the ministry are child nodes of the root. Moreover, for each state office there is a sub tree. An UDK object is assigned to the node representing the organisational unit that is responsible for the corresponding environmental information object.

In order to define a structure tree each department which may use the meta data must be involved. Some of the German states use an organisation-oriented classification. From the public authorities' point of view this classification has a big advantage. During the data updating process a data holder has only to check the UDK objects assigned to the node of the structure tree representing his department in order to guarantee valid data. But searching environmental information may be more difficult because in practice environmental information about the same subject is managed by more than one organisational unit. Additionally, the departments are not forced to work together in collecting meta data. This influences further advantages of meta data management. For instance, data consolidation, identification of redundant data and responsibility problems are not supported in an optimal way. Furthermore, if the organisational structure changes the UDK structure tree has to be adapted, too. Therefore, we propose to use a subject-oriented classification.

2.2 Meta Data Collection

In the literature different strategies were discussed for collecting meta data held by different units of a public authority. If there are no tools available for generating meta data automatically, in (Swoboda 1999) small teams are proposed for collecting meta data based on interviews and database description in a first step. In a second step data holders will have access to the environmental data catalogue with the meta data collected and they can maintain the meta data records. In order to accelerate this process we propose a slightly different strategy. In a first step we identify all data sources (e.g. Web site, publication lists, databases etc.) providing environmental information. A small expert team defines and enters the meta data records. In some cases meta data is already available. In this step data holders are not involved. After this step we start a meta data updating process as described in Fig. 2. It allows the data holders to check the records and to change them. In this step the data holders may also be interviewed by the expert team. The advantage of our strategy is that data holders only work with an environmental data catalogue that provides a lot of meta data already. Data holders are able to see immediately the usefulness of the system and are highly motivated to improve the meta data quality.

3. Data Quality

Ensuring the data quality in information systems is more and more important (Redman 2001). The quality of data is usually described by multiple dimensions, e.g. accuracy, completeness, consistency and timeliness. In order to support the data quality management with respect to the environmental data catalogue we developed several tools that are described in the next subsections.

3.1 Meta Data Updating Process

According to (Rowland 2004) we can consider an UDK object as passive meta data. This means changing an environmental information object that is described by an UDK object, a non-automatic process is needed to update the UDK object. Based on our experiences such a process must be controlled centrally even if a decentralized meta data maintenance strategy is established and data holders are responsible for the updating of the meta data in the environmental data catalogue. Therefore we propose to establish a meta data management group that is responsible for the quality of the UDK objects.

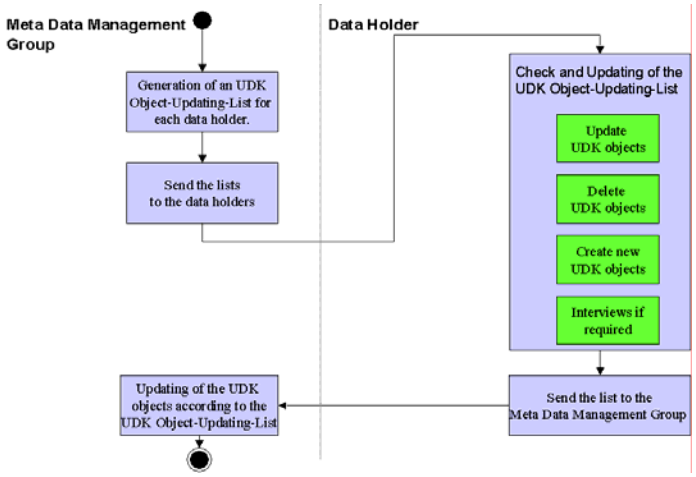

Fig. 2: Meta Data Updating Process.

Fig. 2 shows the main steps of the meta data updating process that is activated by the meta data management group periodically. First of all, for each data holder a list of UDK objects is generated for which the data holder is responsible. This list is called UDK-Object-Updating-List. Based on this list the data holder checks the UDK objects. In case of centralized meta data maintenance data holders change the UDK objects in their lists. Afterwards the lists are sent to the meta data management group responsible for the updating in the environmental data catalogue. In case of decentralized meta data maintenance the data holders directly update the UDK objects in the environmental data catalogue.

3.2 Extension of the Data Model

In order to support the meta data updating process as described in the previous section we propose to extend the data model by defining additional attributes for UDK objects. These attributes can be used for supporting and controlling the meta data updating. Fig. 3 gives some examples.

Attribute	**Description**
- Processing state - Quality level - Comment - Date of next quality check	Information for controlling the updating process. The data holder responsible for an object can use these attributes in order to classify quality and maintenance status of an UDK object.
- Meta Data Source	Internal information about the original source of the meta data record (e.g. extracted manually from …).

Fig. 3: Extension of the data model of an UDK object.

3.3 UDK Object-Updating-List

As described in section 3.1 all UDK objects have to be checked periodically. In doing so data holders have to be informed about critical UDK objects they are responsible for. We developed a tool that generates an UDK-Object-Updating-List as word document for each department. The list is sent to the responsible data holder. The list contains all UDK objects that have to be checked by a data holder. A short description for each UDK object is given. The data holder can access meta data record based on an embedded hyperlink in the word document. In case of centralized meta data maintenance the data holder can enter changed meta data in a field of the word document. The meta data management group updates the corresponding UDK objects in the system. In case of decentralized meta data maintenance data holders updates the UDK objects directly.

3.4 Structure Tree Export Tool

A data holder responsible for updating UDK objects wishes to see the whole structure tree of the UDK for inserting their objects and for checking the structure tree. The Structure Tree Export Tool generates the UDK structure tree as defined in the database and stores it as MS EXCEL sheet. Data holders can check the whole structure tree and the arrangement of their UDK objects. This tool is of importance in case of subject-oriented classification of UDK objects.

3.5 UDK Object Existence Checker

In some cases the department responsible for the Web site of a public authority can provide a list of environmental information objects published on their Web site with the corresponding uniform resource locator (URL). The UDK Object Existence Checker takes this list and checks whether a corresponding UDK object exists. If it exists the name of the related UDK object(s) will be added to the corresponding row of the list. As result the data holder has a list of environmental information objects that are published at the Web site but not mentioned in the UDK database. In order to improve the data quality the responsible data holders have to add the missing meta data record (UDK object).

3.6 UDK Object URL Checker

This tool checks the links in the UDK database. If the Web server returns an error with respect to an URL, the tool generates a warning with the related UDK object. This tool is useful for maintaining the linkage between the UDK objects and the corresponding environmental information objects published on the Web site of a public authority.

4. Conclusions

The main objective of the environmental data catalogue as meta information system is to ensure that the public as well as employees of public authorities have knowledge about environmental information objects available and are able to access it. The success of such a system is strongly influenced by the quality of meta data provided. The processes and tools described in this paper support the meta data management and help to improve meta data quality. Nevertheless further work is necessary. In many cases the meta data updating process is not fully integrated in processes producing environmental information objects. This process integration must be enforced in order to achieve a higher meta data quality.

Bibliography

Fischer, P., Krieger, R., Rietzke, E. (2002): Outlook-Based Concept for the Population and Updating of a Meta-Information System in Environmental Administration. In: A. Rizzoli & J, Jakeman (Eds.): Integrated Assessment and Decision Support, Proceedings of the 1st Biennal Meeting of the International Environmental Modelling and Software Society, Vol 3, pp 378 - 384, Switzerland.

Koordinierungsstelle UDK (2002): Empfehlungen zur Qualitätssicherung der UDK – Datenbestände. Niedersächsisches Umweltministerium.

Kruse, F. (2005): Meta-Umweltinformationssysteme (UDK und gein®). In: Fischer-Stabel (Hrsg.): Umweltinformationssysteme. Grundlagen – Methoden - Anwendungen. Herbert Wichmann Verlag, Heidelberg.

Redman, T. C. (2001): Data Quality – The Field Guide, Digital Press.

Rowland, Ian (2004): Strategic and Tactical Issues in Meta Data Management. Published in DM Direct Newsletter September 17, 2004 issue. http://www.dmreview.com.

Swoboda, W., Kruse, F., Nikolai, R., Kazakos, W., Nyhuis, D., Rousselle, H. (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.