Automation of Metadata Repository Development with XML Schema - Abstract -

Wassilios Kazakos[1], Alexei Valikov[1], Andreas Schmidt[1], Rainer Lehfeldt[2]

1.1. Abstract

Metadata standards have been discussed since years in the environmental domain. Although the community has agreed that metadata are important for information finding, an globally accepted standard is still far away. Different requirements in different environmental application domains, countries and organisations have led to a huge variety on metadata vocabularies and technologies. Even projects which are reusing existing standards tend to adjust these standards slightly to their own requirements. One consequence is that in many projects new metadata repositories are developed or adapted to the new meta data element sets. In this paper we present a novel approach for the automation of metadata repository development. As input the system requires only a XML Schema describing the metadata vocabulary. Out of this schema the user interface, database schema and the search, retrieval and upload mechanism are generated during a batch process. The user interface layout can be further controlled by templates. The System is developed for the NOKIS project and evaluated by generating a metadata repository for adjusted ISO metadata element sets.

2. Introduction

The development of metadata repositories has a long history, starting with electronic library catalogues and is ongoing. Especially in the environmental domain there has been a lot effort in establishing standards for metadata and tools. One of the most successful standardisation example in Europe for a specific interest group is probably the Umweltdatenkatalog[3], containing metadata about environmental information in Germany and Austria (Swoboda et.al. 1999, Swoboda

[1] FZI Forschungszentrum Informatik, Database Systems Department, Haid-und-Neu-Str. 10-14, D-76137 Karlsruhe,
email: {kazakos|valikov|aschmidt}@fzi.de, http://www.fzi.de/dbs.html

[2] Bundesanstalt für Wasserbau DS Hamburg, Wedelerstrasse 157, D-22559 Hamburg, email: lehfeldt@hamburg.baw.de, http://www.hamburg.baw.de

[3] www.umweltdatenkatalog.de

formation in Germany and Austria (Swoboda et.al. 1999, Swoboda et.al. 2000). But neither this initiative nor other promising initiatives, like the Environmental Markup Language (Arendt H.-K., Günther 1999) or EEA's WebCDS (Kazakos et.al. 1999) have led to an European application domain independent standard. The main reason is probably the heterogeneity in requirements from local to international level as well as the different requirements in different environmental application domains, not to mention the lack of awareness of other standards and the lack of willingness to accept a standard developed in other countries and organisations.

In this paper we focus on the development of metadata repositories, and leave the standardisation discussion aside. Since the core functional requirements for such a repositories are well defined, we believe that more effort should be put in automating the development. This would lead in faster development cycles and easier adaptation to adjusted metadata element sets.

The core functional requirements for such a repository consist of defined metadata element set, a storage for the metadata (usually a database but sometimes also the file system), a query and update mechanism for the metadata and a retrieval engine, allowing metadata in different formats. The main variable parts on these requirements are the used metadata element sets, some layout issues and the database system for storing the metadata records.

In the following sections we present our approach to set up a system fulfilling these core requirements, i.e. setting up a database, the query and upload mechanism and a default presentation in one batch process. The only input to the development process is the metadata element set in XML Schema. Further changes to the user interface can be made if necessary and the database set up can be changed.

2.1. Project Background NOKIS

NOKIS[4] (North See Baltic See Coastal Information System) is a joint project of the KFKI, BfG and BAW with the main objectives to establish a prototype meta data information system for the German North Sea and Baltic Sea coastal regions according to the concept of an open system which permits participation of additional partners at any time. To achieve this, within the project a local prototype meta database will be installed in every participating institution and will be integrated via the Internet. One of the key requirements to the NOKIS system is the possibility to adjust slightly the metadata element set on

[4] http://nokis.baw.de

every local site following regional requirements, by adding, renaming or removing some of the elements. FZI develops the system for the NOKIS consortium reusing parts of the CoastBase[5] concept (Kazakos/Kramer/Schmidt 2000).

2.2. Overview

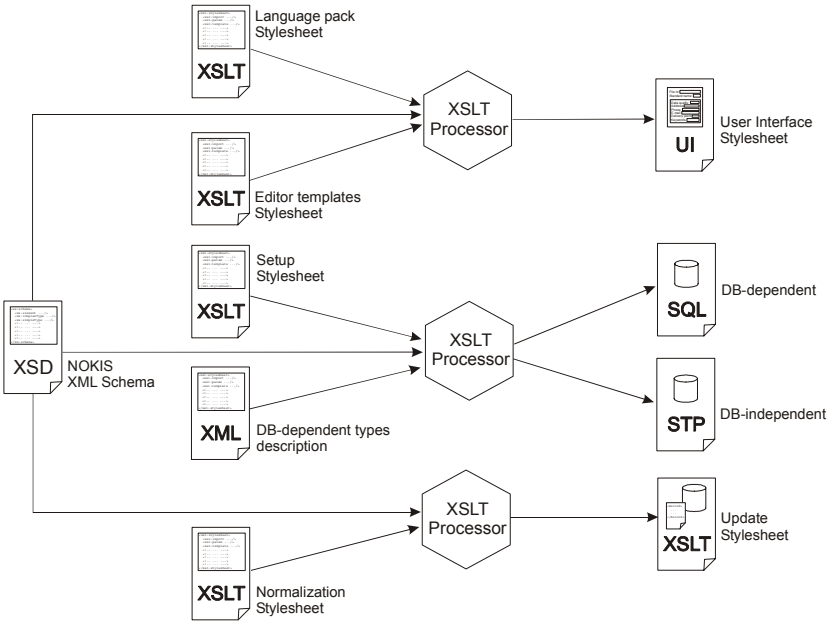
3. Related Work

4. Overall Approach

4.1. Generation Steps

The overall system is developed by following in the main lines an XML-extreme approach as proposed in (Kazakos/Schmidt/Paoli 2001), i.e. using XML-technologies as far as possible, to achieve maximum reusability and easy configuration. We further developed our approach, and set up a system architecture that generates all components of a typical a metadata repository out of an XML schema definition (XSD). The following picture illustrates the overall approach:

[5] http://www.coastbase.org



Firstly, the database schema is generated through a two step transformation process. In the first step a database system independent database set up is generated from of the metadata XSD and a setup stylesheet. In a second step the database system dependent schema is generated out of the database setup and the database system dependent types.

Secondly, the stylesheets for the user interface are generated from the metadata element set XSD by using the editor templates, describing how to present complex and simple XML Schema types. For the editor templates we provide default styles for complex and simple typed elements. In order to support elements with special layout, default templates can be extended (custom UI templates). In order to be able to provide a multilingual user interface, the generation process is additionally controlled by a language pack stylesheet.

Thirdly, in order to be able to store XML documents in the database, we need an update stylesheet, normalizing the XML document and adjusting it to the database schema. This is generated out of the metadata XSD and the normalization templates stylesheet.

4.2. Restrictions

The approach creates out of a given XSD a complete Web-based repository, including automatic database schema and user interface generation. It provides additionally the software modules needed for

the runtime of the system. In order to be able to automate this process, we need to put certain restrictions on initial XML Schema. Firstly, all complex elements must have named complex types defined as sequences of other complex and simple elements. Secondly, there must not be any type cycling (for instance, when type A may have an element of type B and type B may have an element of type A). Thirdly, all simple elements must be either primitive XML Schema types (like string or integer) or enumerations. And finally, attributes must not be used.
Although these restrictions seem to be very burdensome, the set of suitable XML Schemas is still very large. Moreover, the restrictions do not constrain the approach itself, as additional programming effort should allow overcoming most of them.

5. Database Schema Generation

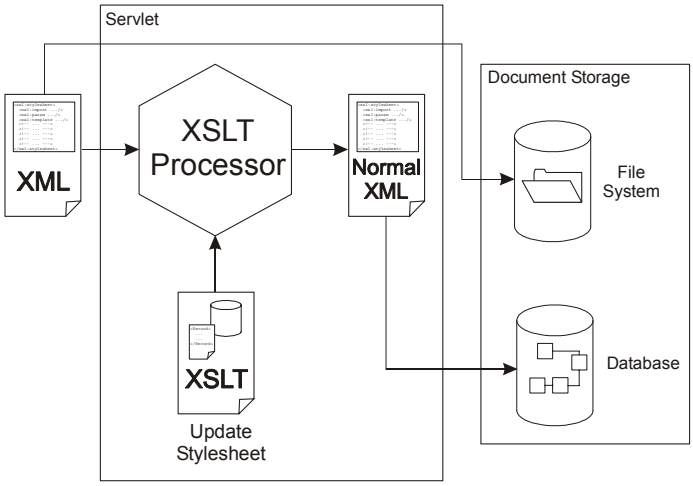
Will be described in the final version

6. User Interface Generation

Will be described in the final version

7. Storing Metadata Records

Will be described in the final version



8. Evaluation

Will be described in the final version

9. Outlook and Further Work

Will be described in the final version

10. Bibliography

Arendt H.-K., Günther (eds.) (1999): Environmental Markup Language (EML). Workshop 1, Berlin.

Kazakos, W., Schmidt, A., Paoli H. (2001): XML based Virtual Catalogue Module in Coastbase. In Lorenz Hilty and Paul Gilgen, editors, Sustainability in the Information Society. 15th International Symposium Informatics for Environmental Protection, volume II, pages 513-520.

Kazakos, W., Kramer, R., Schmidt, A. (2000): Coastbase - The Virtual European Coastal and Marine Data Warehouse. In Armin Cremers and Klaus Greve, editors, Computer Science for Environmental Protection '00. Environmental Information for Planning, Politics and the Public, volume II, pages 646-654.

Kazakos, W., Kramer, R., Nikolai, R., Rolker, C., Bjarnason, S., Jensen, S. (1999): WebCDS - A Java-based Catalogue System for European Environment Data. In Dogac, A., Özsu, M.T., Uluzoy, O (Eds): Current Trends in Data Management Technology, pp. 234 - 249. Idea Group Publisher.

Swoboda, W., Kruse, F., Nikolai, R., Kazakos, W., Nyhuis, D., Rousselle, H. (1999): The UDK Approach: the 4th Generation of the Environmental Data Catalogue for Austrian and German Public Authorities, Proc. IEEE Meta-Data'99, http://computer.org/proceedings/meta/1999/papers/45/wswoboda.html

Swoboda, W., Kruse, F., Legat, R., Nikolai, R., Behrens S. (2000): Harmonisierter Zugang zu Umweltinformationen für Öffentlichkeit, Politik und Planung: Der Umweltdatenkatalog im Einsatz. In Armin Cremers and Klaus Greve, editors, Computer Science for Environmental Protection '00. Environmental Information for Planning, Politics and the Public, volume II Pp 595-607