A METADATA REGISTRY FOR METADATA INTEROPERABILITY

Jian-hui Li *, Jia-xin Gao, Ji-nong Dong, Wei Wu, and Yan-fei Hou

Computer Network Information Center, Chinese Academy of Sciences
**Email:* lijh@sdb.cnic.cn

ABSTRACT

In order to use distributed and heterogeneous scientific databases effectively, semantic heterogeneities have to be detected and resolved. To solve this problem, we propose architecture for managing metadata and metadata schema using a metadata registry. A metadata registry is a place to keep facts about characteristics of data that are necessary for data sharing and exchange in a specific domain. This paper will explore the role of metadata registries and describe some of the experiences of implementing the registry.

Keywords: Metadata, Metadata Registry, Interoperability, Crosswalk, Application Profile

1 INTRODUCTION

Users and applications can easily find, locate, access, and use distributed and heterogeneous scientific databases with the help of metadata. Metadata are especially important for open access to and sharing of scientific data and databases. Different domains, however, will develop or follow different metadata specifications; even the same domain develops different metadata application profiles based on the same specifications according to their special requirements. Consequently, interoperability of metadata is a major issue for scientific data sharing and exchanging.

Metadata Registry is a key solution to solve this problem. The DESIRE (Heery, Gardner, Day, & Patel, 2000), SCHEMAS (UKOLN, SCHEMAS, 2003), and CORES (UKOLN, CORES, 2003) projects are successful examples. Based on the requirements of the scientific databases of the Chinese Academy of Sciences and the Basic Scientific Data Sharing Network, one project of the National Scientific Data Sharing Program, we have designed and developed a metadata registry – the Scientific Database Metadata Registry (SDBMR).

The basic functions of the registry include registering, publishing, and managing schemas. The registry also includes the agencies that maintain and create these resources, application profiles, and element usages, which make up those application profiles. Our registry makes them searchable and also provides services for crosswalking between schemas and creation of application profiles.

In order to describe metadata schemas, we create XML schemas (W3C, 2004) that specify the various attributes of metadata elements. Using Extensible Markup Language (XML) simplifies the metadata schema description encoding process and provides an additional level of integrity checking. The use of XML enables the independent generation of accurately encoded metadata schema definitions.

2 REQUIREMENTS

The project began with an expression of the scope of the work and an initial outline of requirements. An overview of these requirements is presented in Table 1.

Table 1. The requirements of metadata registry

Purposes	To provide information about existing schemas
	To describe the scope and purpose of the schemas
	To identify related schemes
	To describe relevant metadata schemas and their features
Features	Registration: to register metadata schemas into the registry.
	Identification: to identify and distinguish metadata schemas, i.e. title, identifier, version, publication statements, etc.
	Origin: to capture organizations or other agents associated with the metadata schema.
	Description: to capture the purpose, scope of a metadata schema including the types of entities and objects it has been designed to be used for, etc.
	Relationships: to capture relationships between metadata schemas.
	Administration: to administrate the schema registry.

The information within the registry is stored in a form that is both machine-readable and human readable. Users of the registry will typically be *metadata creators* publishing "standard" element sets, *implementers* seeking appropriate schemas or application profiles, *metadata instance creators* seeking guidance on use of schema elements and application profiles, and *researchers* studying schema usage.

Some typical usage scenarios for the SDB Metadata Registry include:

Publishing a description of a metadata schema: An organization provides a resource discovery service for web-based geographical materials that utilizes a metadata schema developed specifically for that purpose. The organization wishes to publish this information to the scientific database community via the registry. The metadata schema can be submitted to the registry and saved in the registry server centrally.

Exploring element usage and creating application profiles: The scientific database organization might mandate use of a simple metadata schema such as the Metadata Schema for Scientific Databases, whilst the geography department might need additional specialized metadata elements for their particular requirements and wish to create a new schema. Using the application profile creation tool, they can easily locate appropriate existing schema elements or be confident in introducing new data elements when necessary.

3 INTEROPERABILITY

Interoperability is the ability of multiple systems with different hardware and software platforms, data structures, and interfaces to exchange data with minimal loss of content and functionality (Bethesda, M.D, 2004). From a methodological point of view (Chan & Zeng, 2006), implementing interoperability may be considered at different levels: schema level, record level, and repository level:

- Schema level – Efforts are focused on the elements of the schemas, being independent of any applications. The results usually appear as derived element sets or encoded schemas, crosswalks, application profiles, and element registries.

- Record level – Efforts are intended to integrate the metadata records through the mapping of the elements according to the semantic meanings of these elements. Common results include converted records and new records resulting from combining values of existing records.
- Repository level – With harvested or integrated records from varying sources, efforts at this level focus on mapping value strings associated with particular elements (e.g., terms associated with subject or format elements). The results enable cross-collection searching.

In the SDBMR Project, our emphasis is on implementing interoperability at Schema level. More than one method will be used.

3.1 Metadata Crosswalking

Crosswalk is used to translate elements in one metadata schema into those in another metadata schema. It is one of the most commonly used methods to implement interoperability between metadata schemas. The mechanism is usually a table that represents the semantic mapping of data elements in one metadata schema (source) to those in another metadata schema (target) based on the similarity of meaning of the elements.

XSLT(the Extensible Stylesheet Language Transformations) (W3C, 2006) provides a powerful implementation of a tree-oriented transformation language for transmuting XML using one vocabulary into either simple text, legacy HTML vocabulary, or XML using any other vocabulary. We use the XSLT language to express the crosswalk tables.

XSLT enables and empowers interoperability. Users, however, can not get background information such as the creator of the crosswalk, creating time, reference to source and target metadata schemas, etc. with single XSLT file. A usable crosswalk should have the following characteristics: a set of mappings between metadata schemas and a well-defined relationship to source and target metadata schemas. The Crosswalks Repository (Godby, Young, & Childress, 2004) uses METS (The Library of Congress, 2006) to model crosswalk. According to our project, we use a database to store this information.

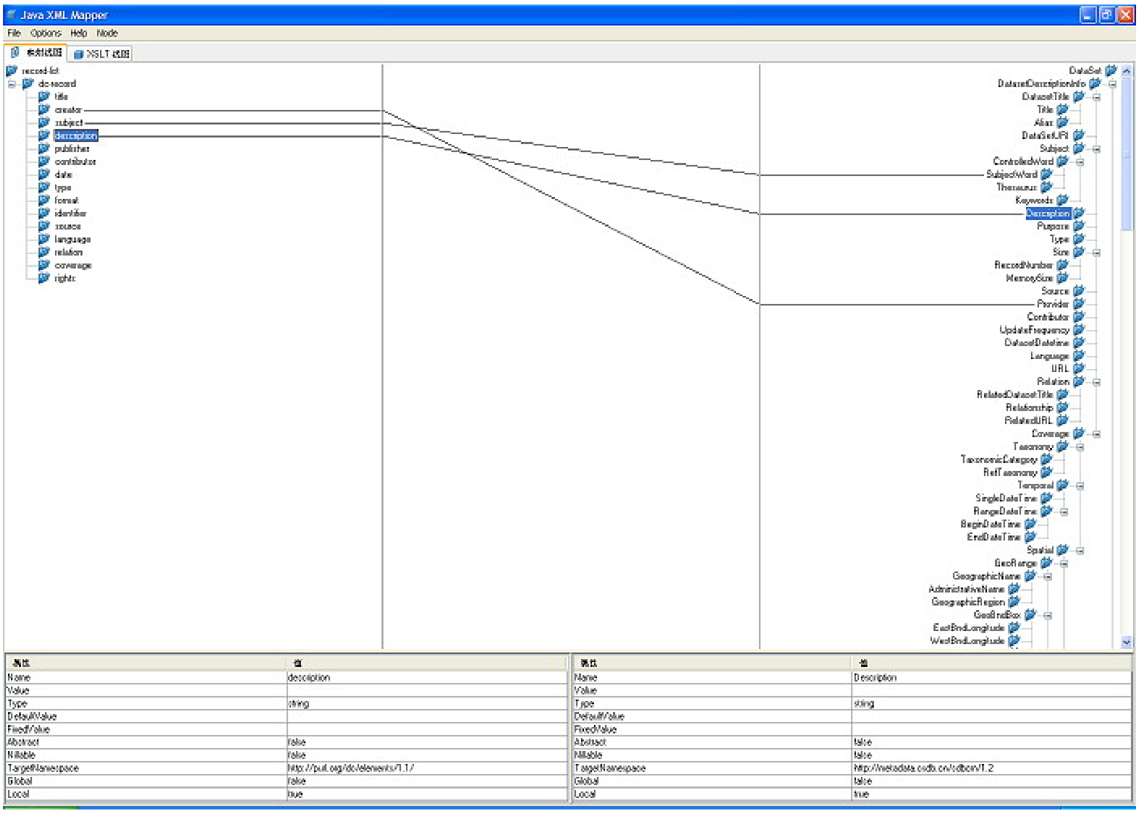

Figure 1. Schema Mapping Tool

To facilitate the creation of crosswalks between metadata schemas, we have developed a graphical schema mapping tool as Figure 1 shows. It provides a friendly interface to help users create relationships between metadata schemas. With this tool, users are able to create proper crosswalks by drag-and-drop and graphical editing without learning XSLT languages.

3.2 Content Conversion

A metadata schema restricts the contents of each metadata element to a particular data type and range of values. Conversions are based not only on defining properties of the source and target metadata elements, but also on the values of those elements. Resolution rules are necessary, for example, between a source and target element using different controlled vocabularies (St. Pierre & LaPlant, 1998). We are now establishing such rules.

3.3 Metadata Re-use

Even within a particular domain, there are special requirements of different users. Oftentimes, there is no metadata schema that meets all needs. Therefore, the concept of "application profiles", the idea that metadata standards are necessarily localised and optimised for specific contexts, are obviously valuable and had been widely adopted (Johnston, 2003).

Application profiles usually consist of multiple metadata elements drawn from one or more metadata schemas, combined into a compound schema, and optimized for a particular requirement. The use of application profiles ensures a similar basic structure and common elements.

An application profile creation tool is being developed to help designers create profiles. Using this tool, designers will be able to know if the terms they need have already been defined or standardized somewhere else and how other projects or services in related areas use those metadata.

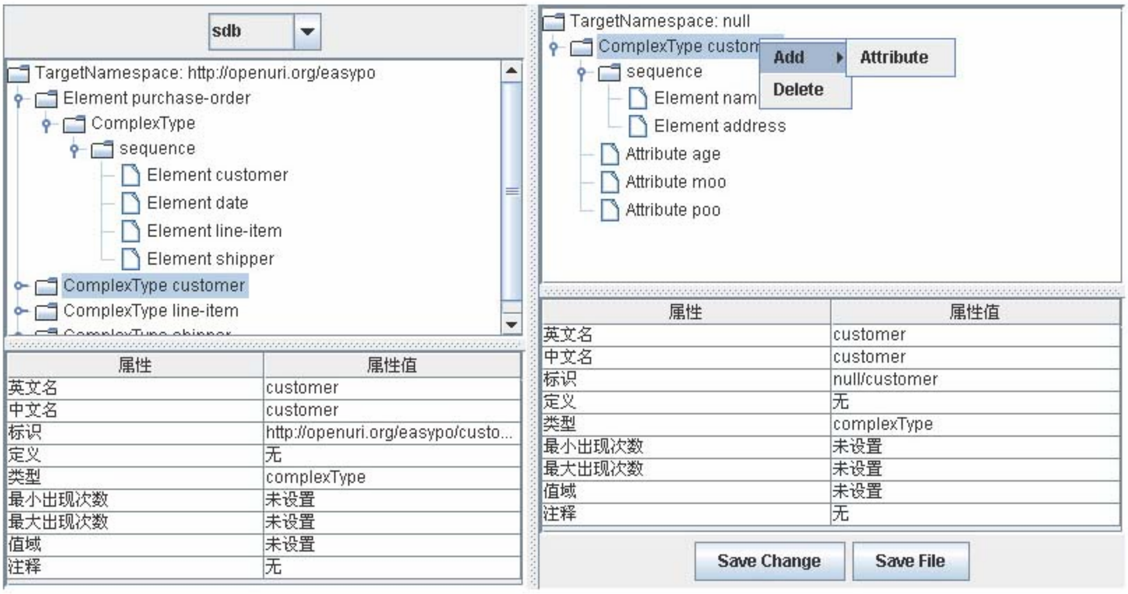

Figure 2. Application Profile Creation Tool

As Figure 2 shows, users can browse or perform keyword searches to find elements and schemes. Results are displayed in the left window. Users can get the information such as usage of elements and can drag and drop appropriate elements into the application profile in the right window. Acquired new schema will be saved locally and then submitted to the registry.

4 CONCLUSIONS

Feedback from interested parties indicates that the services the registry now provides are very useful. Over the next few months we will implement and improve all the functions we have mentioned above. Furthermore, we intend to collaborate with related efforts in order to move forward the registry from a tool for human users to a Web Service.

We will encourage schema creators to register their existing metadata schemas and application profiles. We hope that we will be able to generate commitment to sharing information about metadata element sets.

5 REFERENCES

Bethesda, M.D. (2004) Understanding metadata. Retrieved Nov 20, 2006 from the World Wide Web: http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

Chan, L. & Zeng, M. (2006) Metadata Interoperability and Standardization – A Study of Methodology. Part I: Achieving Interoperability at the Schema Level. *D-Lib Magazine, 12 (6).*

Heery, R., Gardner, T., Day, M., & Patel, M. (2000) DESIRE metadata registry framework. Retrieved from the WWW October, 2006: http://www.desire.org/html/research/deliverables

Johnston, P. (2003) Metadata and interoperability in a complex world. 2003 Dublin Core Conference. Seattle, WA, USA.

St. Pierre, M. & LaPlant, W. (1998) Issues in Crosswalking Content Metadata Standards. NISO White Paper. Retrieved from the WWW October, 2006http:/www.niso.org/press/whitepapers/crsswalk.html

The Library of Congress (2006) METS: Metadata Encoding and Transmission Standard. Retrieved from the WWW October, 2006: http://www.loc.gov/standards/mets

UKOLN, CORES (2003) CORES Registry. Retrieved from the WWW October, 2006:
http://www.cores-eu.net/registry

W3C (2004) XML Schema. Retrieved from the WWW, November 15, 2006: http://www.w3.org/xml/schema

W3C (2006) XSL Transformations (XSLT) Version 2.0. Retrieved from the WWW November 23, 2006: http://www.w3.org/tr/xslt20/