

# XML Technologies And Its Application in Metadata Management System

Zha Xinzheng ,Xiao Yun,Wu Kaichao,Wang Longxiao

Tel: +86-10-62580066-303 Fax: +86-10-62578544

Computer Network Information Center (CNIC) of Chinese Academy of Sciences,  
Beijing 100080, P. R. China

Email: [xzcha@sdb.ac.cn](mailto:xzcha@sdb.ac.cn)

## Abstract

The Scientific Database Project (SDB) has been constructed by Chinese Academy of Sciences (CAS) for more than 20 years. In the past years, about 8.2 TB data resources are available through 313 online databases produced by 45 CAS membership institutes. Faced with so many data resources, in order to exchange them conveniently and make use of them efficiently we developed a system called Metadata Management System. In this paper, authors discuss XML Technologies and data representation ,data modeling and data showing in scientific database. In addition, authors also introduce the processing steps and main function of Metadata Management System.

## 1 Introduction

The Scientific Database Project (SDB) has been constructed by the Chinese Academy of Sciences (CAS) for more than 20 years. In the past years, a lot of data has been collected, classified and stored in different databases. Today, about 8.2 TB data resources are available through 313 online databases produced by 45 institutes of CAS. Faced with so many data resources, we should attach great importance to the problem of how to represent and integrate data formally, in order to exchange them conveniently and make use of them efficiently. We find Metadata and XML technologies are effective solutions to this problem. In this paper, we will introduce the relations of XML Technologies (which include XML, XML Schema and XSL) and metadata representation, metadata modeling, and metadata showing in scientific database. Finally, we also show the application of XML Technologies in the Metadata Management system.

In order to apply metadata theory and xml Technologies to solve the above mentioned question, we developed a system called Metadata Management System. The system is part of system platform of the Scientific Database, with the purpose of developing a metadata management system running on different platforms and combined with the Scientific Database. As the basis of implementing general data platform and accessing through different databases, the software is designed according to Core Metadata Specification for Scientific Database. The system is developed by using Brower/Server architecture and JAVA, JSP/SERVLET, XSLT technologies etc. XML Technologies.

## 2 XML Technologies and its application in Metadata Management System

### 2.1 XML Schema and Metadata modeling

Metadata modeling stands for describing the structure of metadata. It is supported by Core Metadata Specification for Scientific Database (which is defined by CAS) with XML Schema. An XML schema describes the structure of an XML document. In other words, The purpose of an XML Schema is to define the legal building blocks of an XML document. It can define elements (attributes) that can appear in a document, which elements are child elements, the order of child elements, data types for elements and attributes, default and fixed values for elements and attributes. The relationship between XML Schema and XML document is just like a table of a database and records in it.

Let us take a look at one example in the metadata specification that describes the address of a linking man:

```
<xs:element name="contactaddress" minOccurs="0">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="contactaddresstype">
        <xs:attribute name="chinesename" type="xs:string" fixed="contact
          address"/>
        <xs:attribute name="fillable" type="fillabletype"/>
        <xs:attribute name="showable" type="showabletype"/>
        <xs:attribute name="searchable" type="searchabletype"/>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
```

From the example above we can find that using XML Schema to model metadata has some advantages:

Firstly, XML Schema is written in XML, so we don't have to learn another language; we can use the XML editor to edit the Schema files, use the XML parser to parse the Schema files, manipulate the Schema with the XML DOM, etc.

Secondly, XML Schema has Support for Data Types, so it is easier for us to describe permissible document content, to validate the correctness of data, to define data facets (restrictions on data) and to define data patterns (data formats).

Thirdly, XML Schemas are extensible, so we can reuse the Schema in other Schemas, create our own data types derived from standard types, reference multiple schemas from the same document.

### 2.2 XML and Metadata representation

XML stands for the eXtensible Markup Language. Like its predecessor SGML (Standard Generalized Markup Language), XML is a meta-language used to define other languages. However, XML is much simpler and more straightforward than

SGML. As a markup language, XML overcomes the disadvantages of HTML (Hypertext mark-up language) - such as the poor extensibility, the users being unable to define meaningful tags and give them to somebody else to use, the content of a HTML document and its showing style mixed together.

In our developed system, we use XML file to save and represent metadata. But why we use XML not RDBMS(Relational Database Management Systems),I can give some reasons for it. Firstly, the structure of metadata that we defined is complicated, it has many nested relationships and some metadata items allow to be repeated.Secondly, if we use RDBMS to save metadata, then the tables in database cannot reflect the complicated relationships between metadata items, and it is difficult to generate definitions of tables to describe the complicated relationships . In contrast, XML can define the structure of a document in a Self-Describing way. It can be used to describe not only the content but also the structure, thus we can find the relationships between different parts of an XML document easily.

We take a metadata item which describes the address of a linking man as an example:

```
<contactaddress>
  <country>China </country>
  <province>BeiJing </province>
  <city>HaiDian district </city>
  <address>4, 4th,South Street,Zhongguancun</address>
  <postalcode>100080</postalcode>
</contactaddress>
```

From the example above, we can find it appropriate to use XML to represent metadata. XML can be read and written easily and is convenient for exchanging data for different types of applications. This is one advantage of XML. However. Since XML is based on file management, it has deficiencies such as query, parser, security and parallel control. But we can use other technology to avoid these disadvantages.

### 2.3 XSL and Metadata showing

XSL stands for eXtensible Stylesheet Language. As we all know, XML was designed to describe data and to focus on what data is, so there must be something in addition to the XML document that describes how the document should be displayed; and that is XSL. XSL consists of three parts: XSLT is a language for transforming XML documents, XPath is a language for defining parts of an XML document, XSL-FO is a language for formatting XML documents.

XSLT is the most important part of the XSL Standards. It is the part of XSL that is used to transform an XML document into another XML document, or another type of document that is recognized by a browser, like HTML and XHTML. Normally XSLT does this by transforming each XML element into an (X)HTML element. In our developed system, we used XSLT technology to show metadata.

How XSLT works? An XSLT processor (I'll call it an *XSLT engine*) takes two things as input: an XSLT stylesheet to govern the transformation process and an input document called the *source tree*. The output is called the *result tree*. In the XSLT

stylesheet we can define the template which will match the tags in source tree and specifies what to output for building its part of the result tree, and also how to continue processing.

Here is an XSLT stylesheet example:

```
<xsl:template match="contactaddress">
  <table>
  <tr><td>country</td>
  <td>
  <xsl:value-of select="./country"/>
  </td></tr>
  <tr><td> province</td>
  <td>
  <xsl:value-of select="./province"/>
  </td></tr>
  <tr><td>city</td>
  <td>
  <xsl:value-of select="./city"/>
  </td></tr>
  <tr><td>address</td>
  <td>
  <xsl:value-of select="./address"/>
  </td></tr>
  <tr><td>postalcode</td>
  <td>
  <xsl:value-of select="./postalcode"/>
  </td></tr>
  </table>
</xsl:template>
```

In this example, the XSLT stylesheet will transform a metadata item which describes the address of a linking man to a formatted HTML page.

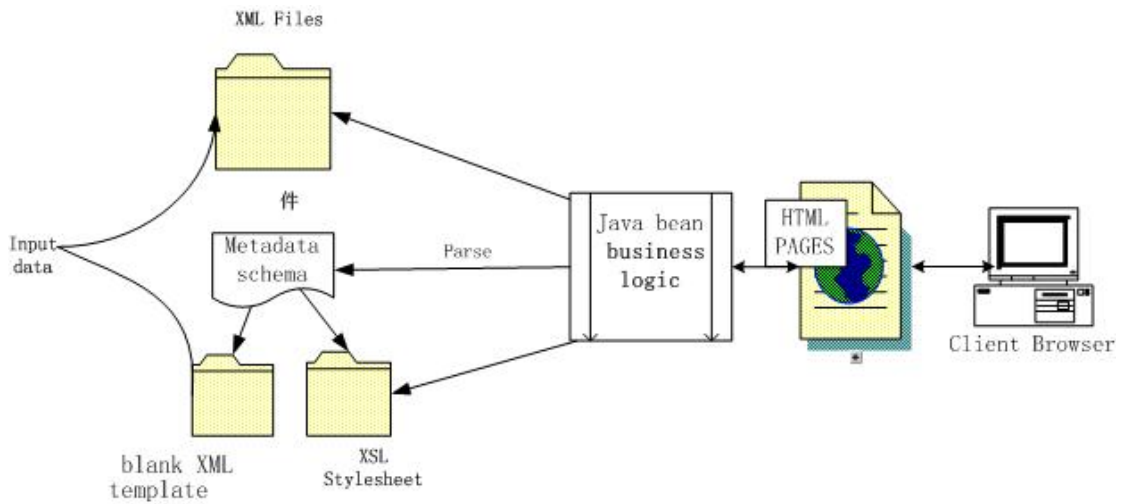
I will give some reasons for using XSLT to show metadata. Firstly, we can save metadata showing style in XSLT files alone. Secondly we need not mix metadata and its showing style together. Thirdly if we want to change the metadata showing style, all we have to do is to modify the XSLT files. Finally, we are able to show the same metadata on different devices (such as PDA, mobile phone etc) or in different applications through defining different XSLT files.

In this way, we can improve the reuse ability of metadata.

#### **2.4 Processing steps and main function of Metadata Management System**

As we mentioned above, we use XML Schema to define the structure of metadata, use Xml document to save metadata, use XSLT stylesheet to save metadata showing style. Since XML Schema, Xml document and XSLT stylesheet are all defined through XML, so we can use the same XML parser to parse them.

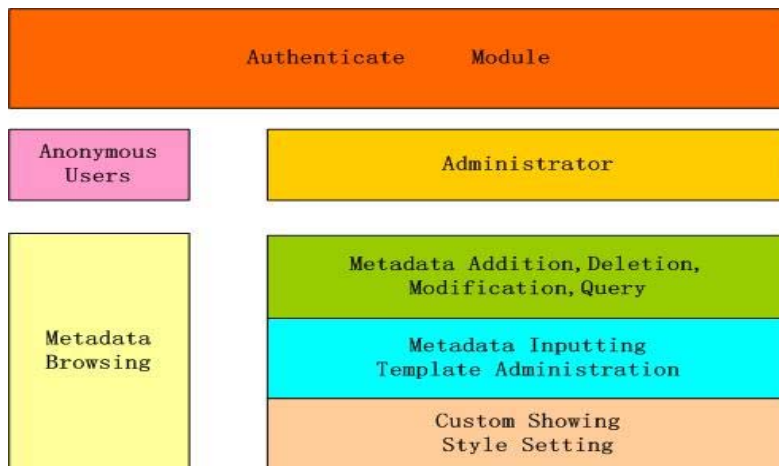
The figure 1 describes the processing steps of the developed system.



**Figure1: processing steps**

From the graph above, we can see that JAVA BEAN business logic will parse the XML Schema to generate blank XML template document and XSLT stylesheet when the software is used for the first time. Then when users input the data of metadata, the system will copy the same blank XML template document and fill it with the input data. In this way, we get and save the metadata in XML files. In addition, JAVA BEAN business logic is responsible for transforming XML documents to HTML pages with the help of XSLT stylesheet.

Our developed system has the main function of metadata addition, deletion, modification and query. Figure 2 shows the main function:



**Figure2: main function**

One question that has to be mentioned here is metadata query. We all know that metadata is saved in XML files, but XML has the disadvantage of poor query. So we find a solution to overcome XML query deficiency in performing metadata query through LDAP Server. LDAP Server solves the problem that we must perform query through some XML files (not a single one).

### **3 Conclusion**

In a word, XML technologies play an import role in metadata representation, metadata modeling and metadata showing. Therefore, we build the developed Metadata Management System with XML technologies. With the development of network and the increasing demand for exchanging data, I think XML technologies is to be widely used in the future.

### **References**

- [1] Erik T. Ray , Learning XML, 2nd Edition , O'Reilly, 2003
- [2] Elliotte Rusty Harold, Processing XML with Java:A Guide to SAX, DOM, JDOM JAXP and TrAX, Addison-Wesley/Pearson,2003
- [3] John Robert Gardner ,XSLT and Xpath:A Guide to XML Transformations ,Pearson Education,2001
- [4] [http://md.sdb.ac.cn/md\\_paper/index.html](http://md.sdb.ac.cn/md_paper/index.html)
- [5] <http://www.w3schools.com/default.asp>