Background

Libraries in Hong Kong, just like their counterparts in other regions where non-roman languages are used, often encounter difficulty in differentiating between authors whose names are transliterated the same way while their vernacular forms are actually different. Lacking vernacular scripts, authority records from LC for these authors provide only limited help. There is clearly a need to overcome the constraints inherent in the existing roman-based and MARC-formatted authority records for non-roman authors, such as Chinese, Japanese, and Korean.

To enhance identification, access and exchange of author information internationally, we believe that a global name repository should be set up in such a way that both vernacular forms and the corresponding roman transliterations for non-roman authors are included, and therefore a more flexible metadata structure will be necessary. XML will be a useful tool for marking up multi-script authority records because it will not only offer greater flexibility in data presentation, but also greater ease in interoperability and exchange of data. The metadata created in XML format will allow more room for manipulation so that the use of name records in the repository can expand from authority control to access facilitation. In the repository only one XML metadata record will be created for a person, and computer programs can do the job of meeting the needs of individual institutions and the requirements of existing cataloging standards such as AACR2.

The purpose of this experiment at HKUST Library is to investigate the advantages of using XML rather than MARC format. We developed a database of some 100 personal name records in XML format and experimented with linking them to the bibliographic records in our library system using the SOAP communication protocol. This project explores an alternative to the model adopted by the HKCAN initiative and is an extension of the propositions presented by Ki-Tat Lam in his paper entitled "XML and global name access control" (OCLC Systems & Services, vol. 18, no. 2, 2002).

Project Goals

- To experiment with creating an XML-based repository of personal name metadata for Chinese authors.
- To collocate the variant name forms of a personal author so that a user can retrieve all of his/her bibliographic works with optimal precision.
- To provide metadata about a person with both non-roman script (Chinese) and corresponding...
To establish a connection between the Repository and our library system, INNOPAC.

Name Access Control Metadata in XML

At this initial stage, we limit the project to only personal name records. We retrieved all our local name (personal) authority records from our database. They were then converted into XML format using the MARC21 XML Schema that was recently released by LC. This schema allows for the retention of fields, sub-fields, indicators, and data in MARC records.

Chinese scripts of different name forms of an entity are added if not already present, including "see from" and "see also" references. The different name forms are then classified with XML attributes.

Owing to limitations imposed by the existing system, we include the existing MARC field labels (such as 100, 400, and 500) in this experiment so that these local records can be easily reloaded to our authority file. In a total XML environment, this will not be necessary. With the flexibility of XML-based data elements, a library can choose to assign any name form to be an established heading in its bibliographic database. All the romanized name forms have their corresponding vernacular scripts (Chinese) included in the metadata.

Technologies, Models and Concepts Adopted

The PERSON Model

In 1998, IFLA issued a final report on FRBR (Functional Requirements for Bibliographic Records), a conceptual framework for bibliographic metadata. FRBR defined the Person entity as the individuals that "are involved in the creation or realization of a work, or are the subject of a work". Although the definition and coverage of Person in FRBR is very brief, the concepts on work, expression and manifestation can apply equally to the Person entity. This study implements and expands this concept as discussed below.

Each record (an XML instance) in the Repository represents a Person. A Person can have multiple Names. A Name is an expression addressing the Person. A Name can be written in many Name Forms, depending on the applicable script, language and romanization schemes. In other words, each of these Name Forms is a manifestation of the Name.

In the Repository, all Name Forms of a Name are linked together by an attribute called Name ID. Each Name Form has an attribute called Reference Type, which allows us to define whether the Name Form is an established heading, or a "see from" or "see also" reference. Another attribute associated with Name Form is Script, which defines the script, language and romanization scheme being used. Figure 1 shows the relationships.

---

Figure 1: the PERSON Model
We believe that by replacing the traditional authority metadata model with the PERSON model, many problems will be resolved, for example:

- The identification problem with romanized names - a one-to-many problem that cannot be resolved without richer links among name forms.

- The multiple "bibliographic identities" problem. According to AACR2, an authority record is created based on bibliographic existence. If an author has more than one identity in bibliographic records, more than one authority record must be created, with one for each identity. This saves work needed for researching different names, but can be very confusing. By using the PERSON Model, one XML instance will only be created for one person. A set of authority records for an author can be generated from this instance as needed.

- Failure of round trip conversion among existing MARC21 models. If you create authority records in the MARC21 Model B format, you will not be able to convert it into Model A. A richer name access control metadata format, such as the PERSON Model, will guarantee the possibility of presenting the data in various existing standard formats on demand.

**MARCXML Schema**

The schema used in the Repository is based on the Library of Congress [MARCXML Schema](http://library.ust.hk/info/reports/xmlnac.html) for MARC21 format. In order to hold the extra data content that is not defined in MARCXML, we have extended this schema to include additional attributes. These attributes are defined separately in a HKUST maintained namespace. It is difficult, if not impossible, to make local modification to the MARC format. However, extension can be easily achieved in an XML environment without breaking the existing schema and affecting data interchangeability.

In this project, three new attributes were added to the `<datafield>` element:
**nameid** - implements the Name ID attribute defined in the PERSON Model.

**script** - implements the Script attribute defined in the PERSON Model

**reftype** - implements the Reference Type attribute defined in the PERSON Model

By using the PERSON Model and the XML format, it is possible to convert name access metadata into authority records, in compliance with existing formats and standards, including MARC21's Model A and Model B. We believe that a well designed metadata format with rich information content will provide us the flexibility for turning the metadata into any standards or formats currently available, or in your imagination.

**SOAP**

We believe that in addition to accessing this Repository from its built-in search interface, it should also be able to be queried using the W3C's [SOAP](https://www.w3.org/TR/soap/) (Simple Object Access Protocol). Z39.50 was a must for allowing database retrieval interoperability in the 1990s, but we believe web services and SOAP is the emerging trend for remotely accessing distributed databases.

We have implemented a small subset of SOAP on this Repository in order to demonstrate the capability of remote authority control from OPAC. Figure 2 shows the messaging between the OPAC and the Repository.

![SOAP Communication between WebOPAC and the Repository](image)

The following are examples of SOAP messages; one is a Request and the other is its Response.

*Request*

*Response*
Innovative's XML MetaServer

HKUST is an alpha test site for Innovative's new XML MetaServer product. We made use of the MetaServer to demonstrate the capability for conducting remote authority control from within INNOPAC.

In a traditional OPAC system, when a user issues an author search, the system will query its authority database to determine the established name forms. If the query is not an established form, the OPAC system will redirect the user to search by the established form. We believe that an OPAC system should offer an alternative to search remotely mounted authority control (or name access control) databases. By doing so, the library will be released from maintaining their own authority database and thus avoid duplication of effort. This concept was supported in this experiment. Figure 3 shows how the traditional and the remote authority control work.

![Figure 3: Traditional and Remote Authority Controls](image)

Tamino XML Server

This repository was built on Software AG's Tamino XML Server. Each PERSON Record is an XML instance of the database. Tamino is the core component of HKUST Library's XML storage and retrieval infrastructure.

Programming Environment

Most of the programs and Web CGI developed for this project were written in PERL. Additional software was used for data transformation and conversion, including LibXML2 for XSLT transformation and MARC4J for MARC formats conversion. Locally developed programs were used to convert character encoding among EACC/ANSEL, INNOPAC brace codes, and UTF-8.
Accessing the Repository

The Repository, which currently contains about 100 persons, can be freely accessed via [http://library.ust.hk/repository.pl](http://library.ust.hk/repository.pl). The experimental OPAC interface, which demonstrates remote authority control, can be accessed at [http://library.ust.hk/xmlopac.pl](http://library.ust.hk/xmlopac.pl).

Conclusion

This project has demonstrated the potential of using XML format for name access control. It allows collocation of various name forms for the same person in a metadata record even if they are in different languages or scripts. It is now, therefore, technologically possible to set up a name access control repository that is useful for and accessible by all libraries in the world. What we need now would be the efforts and contributions from the various concerned parties, such as, national libraries, system vendors, bibliographic agencies, cataloging utilities, etc. And we believe that this repository, once it is set up, will become an immensely useful information resource to libraries and other related institutions in the world.

Posted 2 October 2002
©HKUST Library