Implementation of a KORMARC/EAD integrated system for the Myongji Digital Library Collections

디지털 도서관 콘텐츠 관리를 위한 KORMARC/EAD 통합시스템 구현*

Hyun-hee Kim (김 현희)*

◁ Contents ▷

〈Abstract〉

The study designs and implements a KORMARC/EAD integrated system for the Myongji Digital Library Collections. The purpose of this paper is to design the metadata to Myongji Korean History Collections to provide digital information of high quality to clients, and to develop and implement a model for managing digital library collections. In order to test the model and the quality of the derived metadata, we built a metadata management system, which is connected to the existing KORMARC system. The system consists of two modules: a retrieval and an input module. While in the retrieve mode, one can retrieve KORMARC records of books and archival items, with links to modified EAD files for archival items or to image files for books, in the input mode, one can type two types of data such as a catalog data and an inventory data. Finally, we evaluated the proposed system via mail questionnaires, and propose three suggestions to make this system a much more comprehensive and effective system.

Keywords : EAD, Collection, Ebind, Metadata, Inventory, ISAD(G), USMARC AMC, Integrated system, Archives

〈국문초록〉

　　본 연구는 디지털도서관 콘텐츠를 관리할 KORMARC/EAD 통합시스템을 설계하고 구현하였다. 본 연구의 목적은 명지대학 디지털도서관의 국제한국학 콜렉션과 단행본을 관리할 메타데이터를 설계하고 아울러 기록물정보관리의 모형 구축을 위한 기초 자료를 제시하는 것을 목적으로 한다. 메타데이터를 설계하기 위해서 기록물의 목록 규칙과 관련된 네 가지 종류의 자료 즉 국제표준기록물기술〔ISAD(G)〕, 미국 기록물/필사본 기계가독목록(USMARC AMC), EAD 및 Ebind를 참조하였다.

　　시스템은 검색 모드와 입력 모드로 구성된다. 검색은 KORMARC를 이용하여 단행본과 콜렉션을 통합 검색할 수 있으며 콜렉션인 경우 링크를 통해서 EAD/XML을 활용한 인벤토리로 연결된다. 인벤토리는 기록물에 대한 상세 정보 뿐 아니라 PDF 형식의 전문 이미지 데이터를 제공하며 단행본인 경우 KORMARC의 856 태그를 이용하여 전문 이미지 데이터를 제공한다. 입력 모드는 목록과 인벤토리 정보를 입력할 수 있는 스크린을 제공한다. 시스템에 대한 이용자 만족도, 시스템의 개선점 및 향후 콜렉션을 포함한 다양한 기록물 관리 시스템을 구현하기 위한 기초 자료를 얻기 위하여 설문지를 이용하여 시스템을 평가해 보았다. 평가 분석 결과는 향후 시스템을 갱신할 때 활용할 수 있도록 정리하였고, 제안된 시스템이 좀 더 개선되기 위한 방안 세가지를 끝으로 제시하였다.

　　요어 : KORMARC/EAD 통합시스템, 한국학콜렉션, 메타데이터, 기록물 정보관리

1. Introduction

The development of digital libraries in Korea began in 1996. In 1998, the Myongji University Library (MUL) planned to digitize dissertations and academic articles produced by the professors and graduate students of the Myongji University. However, the MUL changed the plan, because national libraries like the National Assembly Library began to supply full text data of academic papers and dissertations to the public.

The MUL decided to digitize the Korean history collections. So, the collections were to be scanned and to be saved in PDF format, constructing an image database. The collections consist of the LG collection and the KCH (Korean Church History) collection. Myongji library has used KORMARC, a Korean version of MARC, for encoding schemes of traditional books. The KORMARC, like other MARCs, has been pointed for several weaknesses such as lack of link information, inflexibility in formats and compatibility with other metadata. So, the project team planned to implement a KORMARC/EAD integrated system, which has the KORMARC file for cataloging both books and archival items, with a link to a modified EAD file for an archival item or to an image file for a book.

2. Related work

Many libraries abroad, such as Harvard/Radcliffe University Libraries and Minnesota Historical Society Libraries (www.mnhs.org), implemented MARC/EAD integrated systems(Morris 1997). The systems were designed to enter collection-level records onto the library system, and link these system records either to the corresponding inventory files, using the electronic location and access facility provided by MARC field 856. UC Berkeley developed Ebind, which is a method for binding together digital page images using an SGML document type definition (sunsite.Berkeley.edu/Ebind).

In Korea, most libraries are handling archival materials, using very brief catalog formats, just for its preservation, and therefore, specific metadata format and structure like EAD need to be proposed to describe and retrieve them.

3. The Korean history collections

The collections include the LG and KCH collections. The LG collection, collected by financial support of LG, a private enterprise, comprises 8,790 old books and 300 maps. Most of the old books, written by foreigners such as Russians, deal about events and ideas related to Korea.

For example, one of these old books is titled Atlas du Voyage de La Perouse , written by La Perouse, French navigator and published in 1797. This book describes La Perouse's experience of traveling Korean islands and seas. The KCH collection, in addition to 3,300 printed monographs, contains 7,80 pieces, including letters, reports, and maps. The collections are significant resource for historians, historians of religions and sociologists.

4. KORMARC/EAD integrated systems

4.1 Pilot system architecture

Figure 1 provides an overview of the system architecture. As shown in Figure 1, the integrated system consists of two modules: an input module and a retrieval module. The input module allows one to type two types of metadata (a catalog and an inventory) into input screens. While the catalog metadata is required for both books and collections, the inventory one is optional only for collections.

Figure 1. System architecture
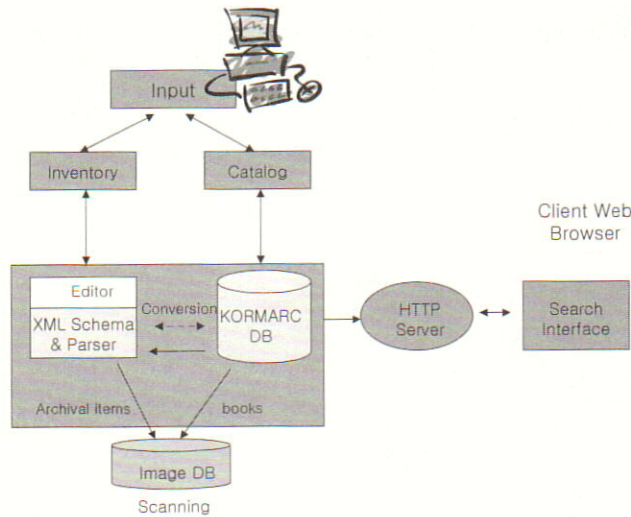
The project team decided to select important collections for EAD markup, a time-consuming job. The inputted inventory elements are automatically converted into an inventory file in an XML format, and the typed catalog ones into a KORMARC file. By using EAD's ENCODINGANALOG attributes that permit the designation of the applicable KORMARC field or subfield, an EAD record is translatable to a KORMARC record, and KORMARC records translated back to EAD ones automatically (Ruth 2001).

The retrieval module allows one to retrieve KORMARC DB stored in MS SQL and display a search result in a KORMARC format. Using the electronic location and access facility provided by KORMARC field 856, one can have access either to its corresponding inventory file, with a link to its image file or to its corresponding image file.

Windows NT server is utilized for the system's server and any type of computer can be used as a client platform. The KORMARC DB was constructed using MS SQL RDBMS and ASP (Active Server program) programs written in Visual Basic Script were created for Web-based searches.

4.2 Database

The proposed system has three kinds of databases: image database, EAD-based inventory database, and catalog database. An overview of the database is shown below in Table 1.

Table 1 Database overview

Database	Indexing	Data type	Software
Catalog DB	Index file, links	KORMARC format	DBMS (MS SQL), ASP
EAD-based inventory DB	Links	XML DTD /XSL format	XML editor, Web browser
Image DB	None	PDF format	Scanning program, Acrobat Reader

4.2.1 Catalog DB

Catalog records are saved in a KORMARC format, which is very similar to a USMARC one, and processed by the OPAC system, called as MULIS (Myongji University Library and Information System). The catalog record has the 856 tag that links one with its inventory file or its image file, which means that inventory and image files are accessed from a catalog gateway. The catalog DB has the index file that includes keywords extracted from the 1XX, 245, 6XX, and 7XX tags of the KORMARC record.

4.2.2 EAD-based inventory DB

The EAD were adopted for its base metadata elements to describe the Korean History Collections. From the EAD element set, we selected core 57 elements, including ⟨dao⟩, ⟨daogrp⟩, ⟨daoloc⟩ and ⟨daodesc⟩ elements (Bouche 1997). At first, we planned to use both EAD and Ebind, a method for binding together digital page images. However, EAD and Ebind have many overlapping elements, so, we decided to use EAD only. For binding together digital page images, we utilized ⟨dao⟩ and ⟨daogrp⟩ elements that are

intended specifically for digital representations from the collection.

4.2.3 Image DB

Collections such as old books and archival materials are being scanned, and saved in a PDF. PDF is selected as an image format, because PDF retains the look of the printed product, and because PDF's Acrobat Reader provides several useful functions such as page-turning and displaying at thumbnail size.
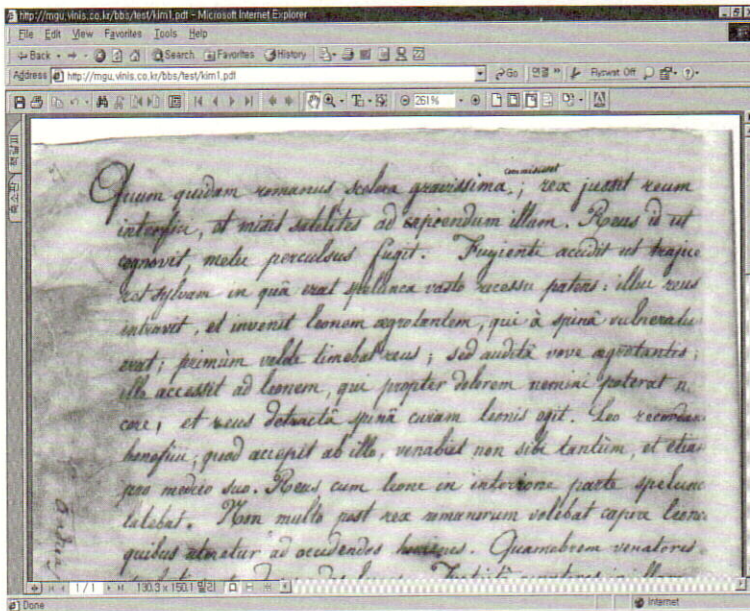
4.3 System implementation

4.3.1 Retrieval module

One can search the catalog database by keyword, creator, and title, or any combination of the three, and also use Boolean operators to connect search terms. When one inputs Kim Dae Kuen , a first Korean priest, as a search term of creator, the full catalog record (Figure 2) that describes about his collection is displayed.

Figure 2 Catalog record

Figure 3 Inventory

Figure 4 XML file

Figure 5 Kim's letter

Kim's collection is divided into four parts: letters, essays, reports and maps. By clicking the link field (856 tag) of the catalog, one can have its inventory record (Figure 3) in an XML DTD/XSL format. The Figure 3 shows the part of the inventory record that includes hyperlinks. By selecting a hyperlinked Link of the inventory file, one can have an access to each item-level image file. Figure 4 displays an xml source file. Figure 5 shows Kim's first letter in a PDF format resulted from clicking the first letter's link. For browsing through the letter, specific computer programs are not necessary, because a PDF Acrobat reader includes functions that enable image files to be displayed page by page or in a thumbnail.

4.3.2 Input module

Figure 6 displays the interface screen where one can input two types of data such as a catalog data and an inventory data. One can input a catalog data in the first part of

the screen, while one can type an inventory data in the latter part of the screen. Typed catalog data is saved and stored in MS SQL and inputted inventory data is stored in XML DTD/XSL files. There are overlapping elements between two files such as creators and titles. So, in order for these elements to be inputted just one time, the system program enables some elements of catalog files to automatically convert into inventory files.

Figure 6 Input screen

4.4 System evaluation

We conducted a questionnaire survey to evaluate the system. We mailed a total of 57 questionnaires to professors, archivists, and librarians and had 30 responses. These surveys were carried out and these survey results were utilized to improve the system. The questionnaire consists of three sections: some background information, system performance, and system usefulness. The survey results indicate that to the system

performance, while seventeen respondents (56.7%) answered   neutral, eleven respondents (36.7%) and two respondents (7%) replied satisfactory and very satisfactory respectively.

Many respondents suggested that some areas need to be improved to make the proposed system become a much more comprehensive and effective system. First, the system's index files that have keywords extracted only from the catalog files should include keywords from inventory files, which allows an integrated retrieval of catalog and inventory files. Second, searching by date range that limits one's search to a specific period should be included. Third, searching inventories directly, without first going through the catalog is necessary for more specific search. Fourth, EAD' full set that consists of 145 elements needs to be utilized to describe various types of archival materials including government documents.

5. Conclusion

In this article, a KORMARC/EAD integrated system for the Myongji Digital Library Collections was designed and implemented. There are some areas where improvements are needed to make the proposed system more efficient and effective. First, Korean rules for archival descriptions like MAD2 (A Manual of Archival Description) need to be developed. Second, MARC21 (KORMARC21) for MULIS should be adopted as soon as possible, to effectively describe various types of materials including archival ones. Third, DOM API should be developed, to effectively handle metadata, because it provides a use interface to browse the tree structure of the metadata descriptions and to manipulate their content.

<Bibliography>

1. Bouche, N. L. 1997. Implementing EAD in the Yale University Library. American Archivist vol.60, 408-419.

2. Dow E. H. 1997. EAD and the Small Repository. American Archivist vol.60, 446-455.

3. Fox, Michael J. and Warren, S. R. (Editor). 1998. Introduction to Archival Organization and Description. Getty Information Institute.

4. Gertz, J. and Stout, L. J. 1989. The MARC Archival and Manuscripts Control (AMC) Format: A New Direction in Cataloging. Cataloging and Classification Quartly, vol.9(4), 5-25.

5. Honhart, F. L. 1988. The Application of Microcomputer-Based Local Systems with the MARC AMC Format. Library Trends. Vol.36, 585-592.

6. Lacy, M. A. 1997. EAD Testing and Implementation at the Library of Congress. American Archivist vol.60, 420-435.

7. Matters, Marion. 1997. Introduction to the USMARC Format for Archival and Manuscripts Control. Chicago: SAA.

8. Morris L. A. 1997. Developing a Cooperative Intra-institutional Approach to EAD Implementation: The Harvard/Radcliffe Digital Finding Aids Project. American Archivist vol.60, 388-407.

9. Ruth, Janice E. 2001. The Development and Structure of the Encoded Archival Description (EAD) Document Type Definition. Journal of Internet Cataloging vol.4 (3/4), 27-59.

10. Ruth, Janice E. 1997. Encoded Archival Description: A Structural Overview. American Archivist vol.60, 310-329.

11. Seaman, D. 1997. Multi-institutional EAD: The University of Virginia's Role in the American Heritage Project. American Archivist vol.60, 436-444.

12. Smiraglia, Richard P. Editor. 1990 Describing Archival Materials: The Use of the MARC AMC Format. New York: The Haworth press.

13. Weber, Lisa B. 1990. Record Formatting: MARC AMC. in Smiraglia, Richard P. Editor. Describing Archival Materials: The Use of the MARC AMC Format.

14. http://europa.eu.int/ISPO/dlm/

15. http://www.gils.net (GILS)

16. http://lcweb.loc.gov/ead/ (EAD)

17. http://www.library.yale.edu/beinecke/manuscript/finding-aids/strc_sgm.htm

18. http://www.archives.ca/

19. http://www.mnhs.org

20. http://sunsite.berkeley.edu/Ebind (Ebind).

21. http://www.cain-rcia.ca

22. http:// www.archives.ca