

기록물정보관리를 위한 메타데이터 설계와 구현

A Study on the Design and Implementation of Metadata for Archival and Manuscripts Control

김현희, 명지대학교 문헌정보학과

정연대, 명지대학교 도서관

안수현, 명지대학교 기록과학 대학원

Kim Hyun-Hee, Jeung Yeon-Dae, Ahn Sue-Hyun
Myongji University

본 연구에서는 기록물 자료를 관리 수 있는 메타데이터 형식을 설계하기 위해서 기록물의 목록 규칙과 관련된 네 가지 종류의 자료 즉 국제표준기록물 기술[ISAD(G)], 미국 기록물/필사본 기계가독목록(USMARC AMC), MARC21 및 EAD(Encoded Archival Description)를 참조한 다음 전통적인 도서관 자료와의 통합검색을 가능하게 하면서 기록물정보의 고유의 속성을 잘 표현할 메타데이터(목록 레코드와 인벤토리)를 설계하여 구현해 보았다.

1. 서론

전통적으로 도서관에서 도서, 비도서 자료를 수집하여 다양한 검색 도구들을 개발하여 관리해 오고 있다. 그러나 기록관이나 일반 도서관에서 수집하고 있는 개인이나 조직의 업무수행 과정에서 생산되는 기록물에 대한 체계적인 관리가 이루어지지 않고 있다. 즉 이제까지는 기록물을 검색/이용한다는 차원 보다는 수집/보존하는 차원에서 주로 관심을 기울여왔다. 본 연구에서는 전통적인 도서관 자료와의 통합검색을 가능하게 하면서 기록물정보의 고유의 속성을 잘 표현할 메타데이터를 설계하여 구현해 보았다.

2. 이론적 배경

2.1 기록물의 목록 규칙

기록물 자료를 관리할 수 있는 메타데이터 형식을 설계하기 위해서 기록물의 목록 규칙과 관련된 네 가지 종류의 자료를 참조하였다. 첫째는 기록물의 목록 규칙인 국제표준기록물 기술[General International Standard Archival Description: ISAD(G)]과 미국 기록물/필사본 기계가독목록(United States MARC for Archives and Manuscripts Control: USMARC AMC)을 살펴본 다음 MARC21을 참조한다. MARC21은 USMARC와 CAN/MARC 형식을 통합한 것으로 기록물 기술에 적합한 요소들을

다수 포함하고 있어서 분석 대상에 포함시켰다. 끝으로 EAD(Encoded Archival Description)에 대해서 살펴본다.

2.1.1 ISAD(G)

국제표준기록물기술은 미국과 캐나다의 기록물기술 원칙을 그대로 유지하면서 서지적 채널을 통해 데이터 교환을 지원하는 것에 그 목적을 두고 있다. ISAD(G)는 7개 영역의 26가지 기술요소를 설명하고 있다(<표 1>).

<표 1> ISAD(G)의 기술요소

기술영역	기술요소
식별사항	참조코드, 제목, 생산일자, 기술수준, 기술단위의 범위와 매체:양,부피,크기
생산배경 사항	생산자명, 행정적/서지적 역사, 기록의 내력, 입수 및 이전의 정보
내용과 조직 사항	범위와 내용, 평가·폐기·계획 정보, 문서의 량, 분류체계
접근과 이용사항	접근 조건, 복제 조건, 문서의 언어/문자, 물리적 특징과 기술적 요구사항, 검색도구
관련 자료 사항	원본의 소재, 복사본의 소재, 연관된 기술단위, 출판 주기
주기 사항	주기
기술 통제 사항	기록전문가 주기, 규칙과 협정, 기술일자

기술요소 중 국제교환 필수 기술정보는 식별사항(참조코드·제목·생산일자·기술수준·기술단위의 범위)의 요소들과 생산자이다.

2.1.2 USMARC AMC

MARC AMC는 기록물과 필사본의 서지통정을 위한 미국 기계가독목록(United States MARC for Archives and Manuscripts Control)을 말하며 1983년에 개발되었다. MARC AMC를 채용하는 대표적인 기관은 북미의 OCLC와 RLIN이다. MARC AMC 레코

드는 다른 많은 MARC 서지 레코드들과 다르다. 가장 명백한 변화는 길이이다. AMC 레코드는 내용, 출처나 맥락, 물리적인 외형, 그리고 자료에의 접근에 관한 정보를 제공하는 다양하고 확장된 주기필드를 가지고 있다. 많은 수의 접근점들 또한 AMC 레코드의 길이에 영향을 주고 있다. <표 2>는 USMARC AMC 포맷의 개요를 나타낸 것이다.

<표 2> USMARC AMC 포맷 개요

TAG	Bibliographic Data
001-008	Control Fields
01X-04X	Number and Code Fields (01X-04X)
05X-09X	Classification and Call Number Fields (05X-09X)
1XX(기본 표목)	Main Entry Fields (1XX)
2XX(서명사항)	Title and Title-Related Fields(24X)
	Edition(25X)
	Publication, Etc.(26X)
3XX(형태기술사항, 조직과 배열)	Physical Description (300)
	Medium (340)
	Organization and Arrangement(351)
5XX(주기사항)	General Note, Restriction on Access(50X-52X)
	Location of Originals/duplicates, Language Note(53X-55X)
	Publications Note, Actions(56X-58X)
	Local Note(59X)
6XX(주제명 부출기입, 색인어)	Subject Added Entry(60X-651)
	Index Term-Genre/form Heading(655)
	Index Term-Occupation(656)
	Index Term-Function(657)
7XX(부출기입)	Added Entry(70X-752)
	Physical Characteristics Access, Host Item Entry(755-773)
85X(소장사항)	Location (851)

<표 3> MARC21 포맷의 개요

TAG	Bibliographic Data
001-008	Control Fields
01X-04X	Number and Code Fields (01X-04X)
05X-08X	Classification and Call Number Fields (05X-08X)
1XX	Main Entry Fields (1XX)
2XX	Title and Title-Related Fields (20X-24X)
	Edition, Imprint, Etc. Fields (25X-28X)
3XX	Physical Description, Etc. Fields (3XX)
4XX	Series Statement Fields (4XX)
5XX	Note Fields (50X-58X)
6XX	Subject Access Fields (6XX)
7XX	Added Entry Fields(70X-75X)
	Linking Entry Fields(76X-78X)
8XX	Series Added Entry Fields (80X-840)
	Series Added Entry Fields (80X-840)
	Holdings, Location, Alternate Graphs, Etc. Fields(841-88X)

2.1.3 MARC21

MARC21은 지금까지 우리가 사용해 왔던 MARC Format의 발전된 한 형태이며 전혀 새로운 개념의 기계가독목록형식은 아니다. 이는 1994년부터 1997년에 이르는 기간동안 USMARC와 CAN/MARC 사용자 위원회에서 두 MARC FORMAT간의 차이를 극복하고 통합된 하나의 FORMAT을 편성하기 위한 노력의 결과 1999년에 USMARC과 CAN/MARC을 통합하여 단일판으로 발표되었고, 2000년 10월 그 개정판이 발표되었다.

<표 4> EAD 기본 요소

헤더 (eadheader)	I.D. (eadid)
	표제 (filedesc/titlestmt/titleproper)
	저자 (filedesc/titlestmt/author)
	날짜 (filedesc/publicationstmt/date)
프론트매터 (frontmatter)	언어 (profiledesc/language)
	번호 (titlepage/num)
	부표제 (titlepage/subtitle)
	헤드 (div/head)
기술 (archdesc)	내용 (div/p)
	헤드 (did/head)
	초록 (did/abstract)
	물리적기술 (did/physdesc)
	소장 장소 (did/physloc)
	전기 (bioghist)
	범위와 내용 (scopecontent)
	조직 (organization)
	노트 (add/note)
	관련 자료 (add/relatedmaterial)
	색인 용어 (controlaccessall)
	접근제한(admininfo/accessrestrict)
	수집 정보 (admininfo/acqinfo)
	자세한 기술 (dsc)

MARC21은 전거용 포맷(Format for Authority Data), 서지정보용 포맷(Format for Bibliographic Data), 분류포맷(Format for Classification Data), 기관·단체정보 포맷(Format for Community Information), 소장자료 포맷(Format for Holdings Data)으로 구성되어 있다. 이 포맷들은 두 개 이상을 연결하여 사용할 수 있다. 예를 들면 전거용 포맷과 서지용 포맷을 결합하여 사용할 수 있고, 역시 서지용 포맷과 소장자료 포맷을 결합하여 사용할 수 있는 점은 USMARC와 크게 다르지 않다. <표 3>은 서지정보용 포맷의 개요를 나타낸 것이다.

2.2 EAD

앞에서 언급한 MARC는 기록물의 검색도구로 사용되기에는 부적합하다. 무엇 보다도 중요한 문제는 MARC는 계층적으로 구조화된 검색도구를 표현하기는 부적합하다. 따라서 적절한 검색도구를 위해서 EAD (Encoded Archival Description) Working Group은 1998년에 EAD(버전 1)를 발표하였다.

EAD의 구조는 세 가지 즉 헤더(eadheader), 프론트매터(frontmatter), 기술(archdesc)로 구분되며 총 135개의 태그로 구성된다. 각 카테고리의 중요한 요소들을 계층별로 열거하면 <표 4>와 같다.

헤더와 프론트매터는 기록물에 대한 개요적인 정보를 나타내고 기술 부분에서 기록물에 대한 상세 정보를 표현하도록 하고 있다. EAD를 표현하는 마크업 언어로는 SGML이 사용되다가 최근에는 XML이 많이 사용되고 있다.

2.3 선행 연구

Yale 대학 도서관(webtext.library.yale.edu)은 기록물 관리를 위해 EAD를 설계하였는데 기존의 135 태그 중에서 핵심적인 44개의 태그를 이용하였다. Harvard/Radcliffe Digital Finding Aids Project와 Berkely, Duke, Stanford 및 Virginia가 함께 진행한 American Heritage Project에서도 기록물관리를 위한 EAD를 설계하였다.

3. 메타 데이터 설계

3.1 목록 레코드 설계

목록 레코드는 USMARC AMC를 참조하여 20개의 필드로 구성된다(<표 5>).

<표 5> 목록 레코드 구조

필드	
1) 참조 코드	11)주제명부출표목 (단체명)
2) 기술 수준	12)주제명부출표목 (회의명)
3) 생산자	13)주제명부출표목 (통일서명)
4) 생산일자	14)주제명부출표목 (일반주제명)
5) 형태기술사항	15)주제명부출표목(지명)
6) 일반주기	16)색인어(장르/형태표목)
7) 요약	17)색인어(직업)
8) 검색도구에 관한 주기	18)색인어(기능)
9) 처리현황정보	19)소장지
10) 주제명부출표목 (개인명)	20)인벤토리의 링크

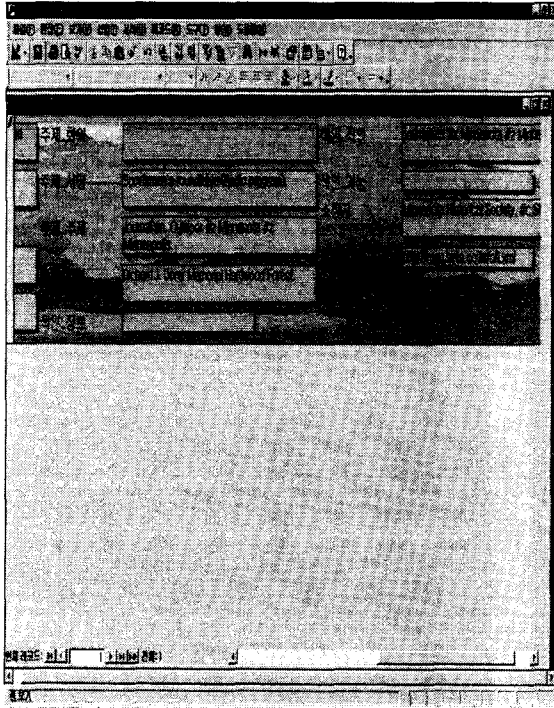
3.2 인벤토리 레코드 설계

EAD의 기본 요소를 참조하여 헤더의 7가지 요소(출판사, 코더 및 언어 추가), 프론트매터의 5가지 요소 및 기술의 13가지 요소들로 구성된 인벤토리 레코드를 설계하였다(<표 4> 참조).

4. 메타데이터 관리 시스템 구현

시스템은 웹 기반으로 구축할 예정이나 현재는 테스트 단계로 목록 레코드는 MS ACCESS로 구축하고 인벤토리 레코드는 XML(DTD)/XSL을 이용하여 구축하였다. 시스템 설계가 끝나면 MS ACCESS 데이터는 MS SQL로 옮기고 ASP(Active Server Page) 프로그램을 이용하여 웹기반 시스템을 구축할

예정이다. <그림 1>은 ACCESS 파일로 구축한 하나의 목록 레코드를 예시한 것이다.



<그림 1>ACCESS 목록 레코드

<그림 2>는 앞의 ACCESS 레코드의 링크 필드를 클릭했을 때 나오는 XML 문서이다.

5. 시스템 평가

웹기반 시스템이 완성되면 두 그룹 즉 기록관리사, 사서 그룹과 이용자 그룹을 대상으로 시스템 만족도에 대한 전통적인 설문지 조사와 웹상의 조사를 하고 그 결과를 시스템 갱신 과정에 활용할 계획이다.

6. 결론

기록물 자료를 관리 수 있는 메타데이터 형

식을 설계하기 위해서 기록물의 목록 규칙과 관련된 네 가지 종류의 자료들과 외국의 기록물 관리 시스템을 참조하여 전통적인 도서관 자료와의 통합검색을 가능하게 하면서 기록물 정보의 고유의 속성을 잘 표현할 메타 데이터(목록 레코드와 인벤토리)를 설계하여 구현해 보았다. 좀 더 체계적인 기록물 정보 관리시스템을 구현하기 위해서는 국내 실정에 맞는 기록물 목록 규칙을 만드는 일이 시급한 것 같다. 이를 위해서는 국제기술표준인 ISAD(G) 검토는 물론이고, 이미 국가기술표준을 채택한 미국의 APPM2와 캐나다의 RAD, 영국의 MAD2(A Manual of Archival description) 등을 충분히 고찰해 봐야 한다.

참고 문헌

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

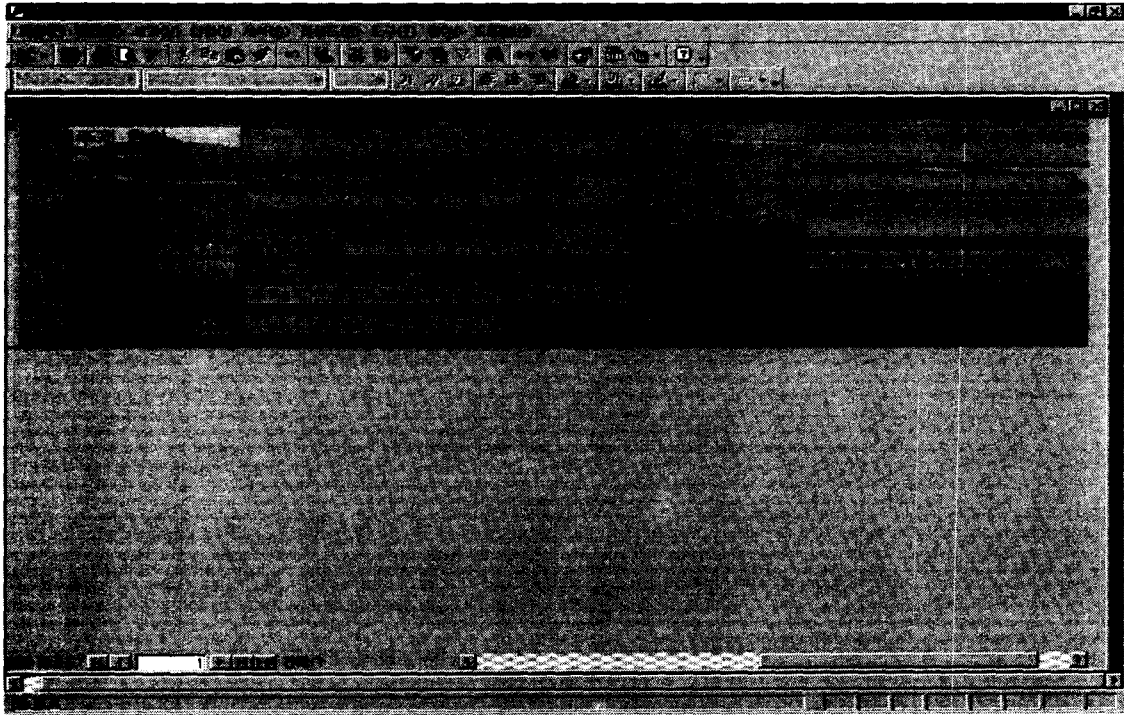
Dow E. H. EAD and the Small Repository. *American Archivist* vol.60, 446-455, 1997

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

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

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

Seaman, D. Multi-institutional EAD: The University of Virginia's Role in the



<그림 2>ACCESS 목록 레코드

American Heritage Project. American Archivist vol.60, 436-444, 1997 <http://lcweb.loc.gov/ead/tglib/tlhome.html>
http://www.library.yale.edu/beinecke/manuscript/finding-aids/strc_sgm.htm
Smiraglia, Richard P. (Editor) Describing Archival Materials: The Use of the Marc Amc Format, 1990. <http://www.archives.ca/>
서혜란. 기록물 기술의 표준화. 기록학 연구, 창간호. 한국국가기록연구원, 2000.
안수현. 한국 기록기술의 표준화 연구. 명지대학교 기록과학대학원 기록관리학과, 제 10회 학술대회, 27-55, 2001.
최정태. 기록학개론. 아세아문화사, 2001.

웹사이트

<http://europa.eu.int/ISPO/dlm/>