

서지정보 기술(記述) 및 교환을 위한 기계가독형 목록 형식



오래전부터 정보자원, 특히 서지정보에 대한 기술 형식으로 널리 이용되어온 MARC은 급변하는 정보환경에 적응하기 위해 노력하고 있다. MARC은 이미 구축되어 있는 대규모의 정보 자원과 포괄성으로 인하여 각종 정보자원에 대한 메타데이터로서 역할을 충분히 수행할 수 있으리라 판단된다.

안계성/ 한국데이터베이스진흥센터 정책연구과 과장

연재 순서

1. 메타데이터의 개요
2. DC(Dublin Core)
3. GILS(Government Information Locator Service)
4. IAFA Templates
5. MARC 이번호
6. PICS(Platform for Internet Content Selection)
7. RFC 1807
8. SOIF
9. TEI header
10. URC(Uniform Resource Characteristics)
11. Warwick Framework RDF(Resource Description Framework)
12. 메타데이터 향후 방향

개요

MARC(Machine Readable Catalog)은 도서관자료의 서지정보를 컴퓨터가 처리할 수 있는 형태로 기록한 기계가독형 목록을 의미하는 것으로 도서관간 레코드의 상호교환을 위해 개발된 것이다.

1965년 미국 의회도서관(LC)의 도서관 자동화 계획에 따라 LC MARC I 이 완성되었고, 이를 수정·보완하여 1968년 LC MARC II가 개발되어 도서관 자동화의 중요한 개념으로 등장하였다. MARC은 초기에는 미국 의회도서관의 정부관련 고문서의 목록을 축적하기 위해 개발된 것으로 국제적 표준형식으로 적용하는 데에는 한계가 따랐다.

따라서 이를 극복하기 위하여 USMARC으로 개정하게 되었고 새롭고 다양한 형태의 정보자료나 통제방식을 지원하도록 확장·개발되어 왔다.

USMARC은 서지레코드 교환 형식 표준인 ISO 2709, ANSI Z39.2 표준 규격을 수용하고 있으며, USMARC 레코드는 ASCII 문자 세트를 사용한다. 따라서 MARC은 국가 서지정보의 축적과 정보처리를 위한 정보교류의 목적으로 다른 도서관이나 기관의 시스템간 연결 매체로의 역할을 하게 되었다.

특히 도서관간의 상호협력을 통한 전산화에 있어서 MARC이 서지 레코드의 유통과 상호교환의 근거가 되어

지금까지도 새로운 자료의 형태를 포함시키기 위한 수정·확장 작업이 계속되고 있다.

각국에서는 초기 LC MARC에 기반하여 국가 MARC을 제정하였는데 영국의 UKMARC, 캐나다의 CANMARC, 일본의 JAPANMARC, 그리고 한국의 KORMARC 등이 해당된다. 그러나 각국에서 다른 형식의 MARC을 제정함에 따라 국가간의 목록 데이터베이스 교환에 어려움이 생기게 되자 국제도서관협회연맹(IFLA)의 주도로 국제적 MARC형식인 UNIMARC이 제정되었다.

MARC 포맷은 도서관 목록레코드의 상호교환과 통일을 위해 개발되었지만 인터넷 기반 환경변화에 대처하여 그 용도가 다양하게 확산될 수 있는 가능성을 가지고 있다.

최근에는 전자문헌을 기술할 수 있는 856필드 '전자정보자원의 소재와 접근방법'을 신설하여 원격접근 컴퓨터파일(인터넷 자원)의 소재 위치와 접근방법을 기술하고 있다. 이 필드에 수록되는 정보는 전자적으로 파일을 전송하거나 서비스를 연결하고 전자저널이나 뉴스레터를 구독하는데 필요한 사항을 기술한다.

특히, 최근에는 다양한 메타데이터가 출현하면서 메타데이터의 표현을 하나의 표준적인 메타데이터로 하자는 주장이 제기되면서 다양한 자원과 메타데이터를 기존 시스템 안에서 통합하여 조직할 수 있는 가장 효율적인 수단으로서 MARC을 사용하자는 논의가 강조되고 있는데, 이는 MARC이 모든 유형의 자원을 수용하여 기술할 수 있고 다양한 접근점을 통하여 이용자의 접근이 용이하다는 점에 근거한 것이다.

MARC 포맷 및 기술내용

USMARC 포맷은 서지데이터(bibliographic data), 전거데이터(authority data), 분류데이터(classification data), 유사정보(비서지자원, community information), 그리고 소장데이터(holdings data)를 위한 5가지 포맷이 제정되어 있고 이중 목록기술 서지레코드 형식인 서지데이터 포맷이 가장 기본이 된다.

현재의 서지데이터 포맷은 1994년판으로서 여기에서는 목록대상자료의 유형에 관계없이 하나의 통합된 형식을 제공하며, 단행본(books), 컴퓨터 파일(computer file), 지도(maps), 악보/녹음자료(music), 연속간행물(serial), 시각자료(visual materials), 혼합자료(mixed materials) 등 각 자료의 형태에 대해 정해진 명세에 따라 레코드를 생산하도록 되어 있다.

MARC을 기존의 인쇄된 카드목록과 비교하여 그 특징을 살펴보면, ①인쇄된 카드목록보다 내용, 규칙, 형식면에서 더욱 정확하고 표준적이며 융통성 있는 서지정보를 유지할 수 있다. ②개별 도서관에서 요구하는 목록정보에 대한 특정사항에 대하여 융통성 있는 조정이 가능하며 각 도서관은 기본 MARC 레코드에 자관의 고유정보로 관내 이용과 소장정보 재배열을 위한 필드 등을 추가할 수 있다. ③OPAC(온라인목록)의 열람서비스나 목록카드의 제작시에 통제된 표준적 어휘와 구조를 제공한다. ④MARC 포맷을 기준

으로 공동편목이 가능하며 호환시스템의 경우에는 전체를 바꿀 필요없이 부분적으로 수정할 수 있어 작업이 수월하다. ⑤공유목록의 질적 수준의 상승 효과와 편목작업의 경비 및 시간 절약에 따른 경제성을 유지할 수 있고 정보접근과 제공을 위한 여러 방식에 대한 선택의 폭을 넓힐 수 있다.

■ MARC 레코드의 구조

MARC 레코드는 레코드의 기본 구조(record structure), 레코드 데이터 내용(data content of the record), 그리고 내용지시기호 혹은 태깅 방법(content designation 혹은 tagging schema)의 세 개의 기본요소로 구성된다. USMARC의 레코드 포맷은 <그림 1>과 같다.

리더(Leader)는 레코드의 첫 번째 필드로 24자 길이의 고정장 필드로 레코드를 식별하고, 레코드의 처리에 필요한 사항을 부호로 기술한다. 리더는 내용지시기호인 태그(Tag), 지시기호(Indicator), 하위필드코드(Subfield code)를 사용하지 않는다.

리더가 어떤 레코드인지를 기록하는 정보 필드인 반면, 레코드 디렉토리는 책의 목차와 같은 것으로 데이터필드부에 속하는 각 데이터필드의 길이와 위치를 지시해 주는 정보를 포함한다. 각 디렉토리 항목의 길이는 3자리 태그와 4자리 데이터 필드길이, 그리고 5자의 시작번지를 모두 합한 총 12자리를 갖는다. 이러한 레코드 디렉토리는 자동으로 생성된다.

리더 (Leader)	레코드 디렉토리 (Directory)	제어필드 (Control Field)	데이터필드 (Data Field)
----------------	-------------------------	-------------------------	-----------------------

<그림 1> USMARC 레코드 포맷

〈표 1〉 데이터필드 리스트

태그	필드내용
0XX	제어정보, 번호, 기호 (control information, numbers, codes)
1XX	기본표목 (main entry)
2XX	표제 및 표제관련사항 (titles, edition, imprint)
3XX	물리적 기술 등 (physical description, etc.)
4XX	총서사항 (series statements)
5XX	주기사항 (notes)
6XX	주제명 접근필드 (subject access fields)
7XX	부출표목:연결표목(name, etc. added entries or series:linking)
8XX	총서명 부출표목(series added entries:holdings and locations)
9XX	자관용 필드(reserved for local implementation)

〈표 2〉 세부 데이터필드 내용

태그	필드내용
X00	개인명 (personal names)
X10	단체명 (corporate names)
X11	회의명 (meeting names)
X30	통일서명 (uniform titles)
X40	서지적 표제 (bibliographic titles)
X50	주제어 (topical terms)
X51	지리명 (geographic names)

제어필드는 '00'으로 시작되는 태그를 가지고 001에서 009까지의 필드를 사용한다. 지시기호와 하위필드코드를 사용하지 않으며, 각 필드는 필드종료기호로 끝난다는 특징이 있다. 데이터필드는 데이터 자체가 수록되는 필드로서, 제어필드와 함께 가변장으로 이용한다. 데이터필드내에서는 지시기호와 하위필드코드가 모두 사용된다.

〈표 3〉 856 필드의 하위필드 코드

하위필드코드	내용	하위필드코드	내용
\$a	호스트명	\$n	하부필드 \$a의 호스트 소재지명
\$b	IP주소	\$o	운영체제(OS)
\$c	파일압축 정보	\$p	포트
\$d	파일경로	\$q	파일 전송방식
\$f	파일명	\$s	파일 크기
\$g	마지막 파일명	\$t	단말기 에뮬레이션
\$h	서비스 요청 처리기	\$u	URL
\$i	지시사항	\$x	비공개 주기
\$k	패스워드	\$z	공개 주기
\$l	로그온/로그인	\$2	접근 소스
\$m	접근지원제공자	\$3	지정 자료

자정보자원의 소재와 검색은 물론 관련된 서비스와의 연결이 가능하다. 856 필드에는 URL을 데이터로 사용하거나, 기타 독립된 하위필드에 필요한 식별정보를 표현할 수 있다. 856 필드의 하위필드코드는 〈표 3〉과 같다.

■ 레코드의 기술규칙

MARC 포맷은 시스템간의 레코드 교환을 위한 형식이기 때문에 개별 시스템에서 사용되는 내부 형식과는 무관하며, 내용을 기술하는 규칙과 직접 관련되는 것은 국제표준서지기술법(ISBD)이며 USMARC의 경우에는 AACR2도 관련된다.

ISBD는 MARC에 사용되는 서지요소와 형식과 순서, 구두점 등 기술의 내용과 순차를 표준화한 것으로, 기계가독형 목록에 큰 영향을 주었다. 이러한 규칙의 이용은 서지 포맷에 맞게 일관성 있는 기술을 유지하기 위함이다. 일관성 있는 레코드의 기술은 상호교환을 가능하게 하며 정확하고 양질의 메타데이터를 생성하는데 도움이 된다.

실례로, 기술규칙에서는 표제(title)로 선택될 수 있는 주정보원을 설정해주고 이를 정확히 선정하기 위한 각종 규칙이 정해져 있다. 따라서 이 규칙에 맞추어 정확히 입력한다면 누가 해당 정보원에 대한 MARC 레코드를 생성하더라도 똑같은 표제가 만들어지게 되는 것이다.

MARC과 관련된 논의사항

MARC의 사용환경이 계속적으로 바뀌면서 MARC와 관련된 많은 논의가 이루어지고 있다. 특히 국제표준으로서 MARC를 사용할 경우에 나타나는 다국

어 처리 문제, 레코드간의 연결 문제, 그리고 네트워크 환경에 있어서 프로토콜 개발 등이 큰 이슈가 되고 있다.

■ 다국어 문제

USMARC은 다국어 처리를 위해 기술적 표준으로서 ASCII와 ANSEL (American National Standard for Extended Latin : Z39.47) 문자세트를 지원하고 있고 초기 USMARC에서는 비로마자에 대해 이를 로마자로 표현하기 위한 여러 가지 표준을 적용시켜 왔다.

그러나 비로마자 대체 형식의 연결을 위해서는 태그 880을 갖는 대체 형식 필드를 반복적으로 사용하여 비로마자 텍스트를 포함할 수 있도록 한다. 그 대표적인 예는 RLG(Research Libraries Group)와 LC가 공동으로 개발한 CJK(Chinese, Japanese and Korean) 문자세트를 들 수 있으며 이는 레코드내의 비로마자 텍스트를 로마자로 표기된 필드와 비로마자의 필드를 연결하는 방법으로 생산하는 것이다.

다른 메타데이터가 영문으로된 자료를 대상으로 개발된 것에 비해 확장된 USMARC은 비로마자 정보자원까지도 처리 대상으로 포함하기 위해 구체적인 연구가 수행되고 있다.

■ 레코드간의 연결 문제

MARC은 한 레코드와 다른 레코드를 직접 연결할 수 있을 뿐만 아니라 특정 필드를 사용하여 서지적으로 관련된 레코드도 연결할 수 있다.

연결표목 필드에 의한 레코드 연결(76X-77X)에서는 계층관계 표현을 위해 770, 772 필드(부록이나 보유판, 특별호

〈예〉 USMARC 레코드 기술 사례

```
LDR *****nkD##2200000#a#4500
001 dag96000488/PP
005 19960126155133.0
007 kh|bh|
008 940000q18561860xxunnn |||||###kn|||
037 $aLC-USZ6-1926$bDLC$c(b&w film copy neg. post-1992)
050 00$aDAG$bno. 1005$u(P&P) USE SURROGATE
040 $aDLC$bDLC$egihc
245 00$a(Edward Morris, three-quarters length portrait, full face, seated at desk)$h(graphic).
260 $c(between 1856 and 1860)
300 $a1 photograph :$bsixth plate daguerreotype.
530 $aAvailable on the Internet:$bLibrary of Congress World Wide Web site (http://lcweb.loc.gov).
500 $aPhotographer unidentified.
500 $aWritten on back of plate: Ed. Morris, Senator from Sinoe County.
500 $aHallmark: Rinhart 20a.
500 $aWas part of LOT 8554.
506 $aOriginal served by appointment only.
541 $cTransfer:$aManuscript Division.
580 $aForms part of the American Colonization Society Records, 1792-1964 (Library of Congress).
580 $aForms part of the Daguerreotype Collection (Library of Congress).
545 $aThe American Colonization Society was organized in 1817 to resettle Afro-Americans in Liberia.
600 10$aMorris, Edward,$db. 1802.
655 #7$aPortrait photographs$y1850-1860.$2gmgpc
655 #7$aDaguerreotypes$y1850-1860.$2gmgpc
773 1#$tDaguerreotype Collection (Library of Congress)$w(DLC)##95861318
773 1#$aAmerican Colonization Society.$tRecords.$d1792-1964.$w(DLC)##mm7810660
856 7#$3b&w film copy neg.$ddag$f3d01926$2file
852 $aLibrary of Congress$bPrints and Photographs Division$aWashington, D.C. 20540-4840
USA$ndcu
985 $app/dag
```

와 그 모체가 되는 자원과의 관계), 773 필드(단행본의 구성요소와 단행본과의 관계), 760, 762 필드(연속간행물에서 총서와 하위총서와의 관계)를 사용한다.

또 연관관계 표현을 위해서는 765, 767 필드(원서와 번역서와의 관계), 775 필드(특정자료와 상이한 판간의 관계), 776 필드(물리적으로 상이한 형태로 간행된 자료간의 관계)를 사용하며, 전후관계를 위해 780 필드와 785 필드(선행저록과 후속저록간의 관계)를 사용하고 있다.

연관저록 복합주기 필드에 의한 레코드 연결(580 필드)은 다른 레코드와의 복합

적인 관계를 표현하는 주기로 연결표목 필드에 기술된 정보자원에 종속된 경우로만 한정하는데 580 필드에서 표현하는 관계는 공식적으로 간행된 복제물로서 원본의 내용을 상세하게 담고 있는 자료, 연속간행물 이외의 자료나 단행본과 관련을 가진 연속간행물, 두 개 이상의 서로 다른 간기로 간행되어 두 개 이상의 레코드를 필요로 하는 연속간행물 등이 해당된다.

기타 방법으로는 제어번호/표준번호에 의한 서지적 관계 표현, 총서명 부출표목에 의한 서지적 관계 표현, 주기필드에 의한 서지적 관계 표현, 형태사항을 통한 특

(표 4) DC와 MARC의 매핑

Dublin Core element	USMARC
Title	245 \$a title statement
Creator	100 \$a main entry-personal name 110 \$a main entry-corporate name
Subject	653 \$a index term-uncontrolled 650 \$a subject added entry-topical term 050 \$a LCCN(Library of Congress Call Number) 082 \$a DDCN(Dewey Decimal Call Number)
Description	520 \$a summary
Publisher	260 \$a publication, distribution, etc (Imprint)
Contributors	700 \$a added entry-personal name 710 \$a added entry-corporate name 720 \$a added entry-uncontrolled name
Date	260 \$c date of publication, distribution
Type	516 \$a type of computer file or data note
Format	538 \$a system details note Identifier 024 \$a Other standard 856 \$u URL 020 \$a ISBN 022 \$a ISSN
Source	786 \$t data source entry
Language	041 \$a language code 546 \$a language note
Relation	787 \$n nonspecific relationship entry/note
Coverage	500 \$a general Note Spatial: 255 \$c cartographic mathematical data 034 \$a coded cartographic mathematical data Temporal: 513 \$b type of report and period covered note 045 \$a time period of content
Rights	506 \$a restrictions on access note

정자료와 딸림자료와의 관계(300 필드), 서지 포맷에서 URL을 사용하여 자원과 서지레코드간의 연결(856 필드) 등으로 자원간의 관계를 표현하고 있다.

■ 프로토콜 문제

인터넷에서 서지 정보를 탐색하고 검색하기 위한 Z39.50 프로토콜은 원래 MARC 레코드의 탐색과 검색의 편의를 도모하기 위해 설계된 것이다. 이 프로토콜은 MARC 레코드 데이터베이스를 Z39.50 클라이언트에서 서버를 통해 탐색하는데 사용할 수 있다.

이때 검색된 레코드는 MARC 레코드 형식을 취하게 된다. Z39.50 프로토콜은 클라이언트를 통해 입력된 탐색어를 서버가 인식할 수 있도록 속성집합을 사용하며 이 속성집합의 표준으로서 bib-1 속성집합이 정의되어 있다. 결국 bib-1 속성집합은 MARC와 같은 서지 레코드와의 매핑을 가능하게 해 준다.

USMARC의 기술 사례

USMARC를 근거로 각국에서는 자국의 자료처리에 가장 적합한 MARC를 개발하였고, MARC의 국제적인 교환을 위

해 IFLA에서는 UNIMARC를 제정하였다. 그리고 실제 도서관마다 USMARC를 사용할 경우 일부 내용을 확장하기도 하였다.

이렇게 확장된 포맷(extended format)을 슈퍼세트(supersets)라 하는데, OCLC MARC, RLIN MARC, Utlas MARC과 같은 것이 USMARC를 확장한 대표적인 예이다. 확장된 포맷은 순수 포맷(pure format)으로는 처리할 수 없는 기능을 제공하여 자원의 처리요구를 만족시켜주게 된다. 다음은 USMARC 레코드의 기술 사례이다.

다른 메타데이터와의 관계

메타데이터의 기본 사상은 형식과 목적을 단순화하여 네트워크 자원의 저자가 직접 메타데이터를 기술하는 것이지만 실제로 MARC과 같은 기술 수준으로 메타데이터를 유지하기 위해서는 많은 비용과 노력이 필요하다.

따라서 MARC과 독립된 다양한 메타데이터로 디지털정보자원을 기술하여 MARC으로 구축된 방대한 서지 데이터와 공존하게 하는 방안이 제시되고 있다. 결국 다양한 메타데이터와 MARC 포맷간의 데이터 변환 수단의 개발이 중요하게 대두되는데 실제로 이러한 상이한 형식간의 데이터변환을 위한 메타데이터간의 매핑이 시도되고 있고, 일부 데이터 요소는 확장하거나 수정할 필요가 생겨나게 되었다.

대표적인 예로서 DC(Dublin Core)와 MARC과의 매핑은 다음과 같이 이루어질 수 있다.

결론 및 향후 방향

최근 논의되고 있는 모든 유형의 메타데이터 통합은 두가지 방향으로 전개되고 있는데, MARC로 통합하자는 것과 다른 메타데이터로 통합하자는 것으로 대별된다.


MARC로 통합하자는 주장을 지지하는 이유는, ①포맷 자체가 국제 표준으로 제정되어 있고, ②지구상의 어떠한 형태의 자료라 하더라도 형태나 내용면에서 완벽하게 표현이 가능하며 ③새로운 유형의 자료를 표현하기 위한 개정·보완작업이 계속해서 이루어지고 있고 ④표준 기술규칙을 따름으로 인해 정확한 메타데이터를 생성할 수 있다. 또 ⑤현재 대량의 데이터 베이스가 구축되어 있으며 ⑥MARC 포맷으로 자동화가 되어 있는 도서관이나

정보전문가와 같은 중개인에 의해 여과된 정보를 이용자에게 제공할 수 있다는 점을 들 수 있다.

MARC으로의 통합을 반대하는 의견은, ①기술규칙 적용이 까다로워 메타데이터 생성이 어렵기 때문에 메타데이터를 생성하기 위한 전문가나 중개인이 필요하며 이는 원자료의 생성의도와 다른 메타데이터 생성이 가능하다. 또한 ②자원간의 관계 즉, 자원의 계층적 구조를 세밀하게 표현할 수 없다. 그리고 ③기본표목(1XX)의 설정은 검색에 단점이 될 수 있다는 점을 강조한다.

이와 같은 단점을 인식하여 보완·확장해 나간다면 MARC은 가장 전문적인 메타데이터로 통합되거나 최상위 메타데이터로서 그 존재 위치가 굳어질 수 있다.

그리고 실제로 현재 자동화된 도서관시스템에서는 내부적으로는 다른 포맷으로 저장되어 있을지라도 MARC 포맷으로 입력하여 검색하는 것이 대부분이다. 특히 856 필드의 변화는 OPAC의 소프트웨어에 절대적인 영향을 끼치게 될 것이다.

이러한 상황에서 앞으로는 통합보다는 중개자에 의해 제어될 수 있는 전문적 메타데이터로서 MARC이 개선·확장되어야 할 것이다. 오랜 시간을 거치며 축적되어 온 MARC 체제를 무시하고 새로운 체제로 재정비한다는 것은 정보량의 규모와 투자될 자원을 고려해 볼 때 매우 어려운 일로 판단되며, 따라서 MARC의 확장·개정을 통해 전문 메타데이터로서 완벽한 체계를 갖추도록 하는 방향으로 나아가야 할 것이다. 

참고 문헌

- 김태수. 1997. 목록의 이해. (미정고판). 서울:연세대학교 문헌정보학과.
문헌정보처리연구회. 1998. 메타데이터의 형식과 구조. 서울:문헌정보처리 연구회.
문헌정보처리연구회. 1996. 기계가독목록의 이해. 서울:문헌정보처리 연구회.
안계성. 1998. 디지털 정보자원의 효율적인 표현을 위한 메타데이터.
KISEC Working Paper. 서울:한국정보검색위원회.
정영미. 1997. 지식구조론. 서울:한국도서관협회.
A review of metadata : a survey of current resource description formats.
http://www.ukoln.ac.uk/metadata/desire/overview/rev_14.htm
Dublin Core/MARC/GILS Crosswalk.
<http://lcweb.loc.gov/marc/dccross.html>
Expanded definition of USMARC Leader/06 type of record.
<http://lcweb.loc.gov/marc/leader06.html>
Guidelines for the use of field 856. rev. August 1997
<http://lcweb.loc.gov/marc/856guide.html>
The USMARC Formats : background and principle.
<http://lcweb.loc.gov/marc/96principl.html>
USMARC Foramt for bibliographic data : field list. update No.3(July 1997)
<http://lcweb.loc.gov/marc/bibliographic/ecbdlst.html>