전자정보자원의 목록 필요성에 관한 연구*

A Study on the Necessity to Catalog the Digital Information Resources

도 태 현(Tae-Hyeon Doh)**

〈 목 차 〉

Ⅰ. 시작하는 글
Ⅱ. 전자정보자원의 새로운 서지 도구들
Ⅲ. 네트워크 세계와 메타데이터의 빛과 그늘
1. 전자정보자원의 특성과 도서관의 역할
2. 검색엔진과 메타데이터의 문제점
Ⅳ. 도서관 목록에 대한 재인식
Ⅴ. 전자정보자원 목록의 나아갈 방향
Ⅵ. 마치는 글

초 록

이 연구는 도서관에서 네트워크 상의 정보자원을 포함하는 전자정보자원의 서지 도구로서 인터넷 검색엔진이나 메타데이터가 갖는 문제점과 전통적 목록의 필요성을 다루었다. 검색엔진이나 메타데이터는 전자정보자원을 신속하게 색인하거나 조직할 수 있게 한다. 그러나 이와 같은 신속성 외에도 목록과 같은 전통적 도서관의 서지 도구들은 기술의 정확성, 완전성, 표준성 등이 절실히 요구되는 것이다. 그러므로 도서관의 관점에서는 전자정보자원에 대하여도 완전한 목록을 작성할 필요가 있다. 목록의 복잡성과 이로 인한 고비용 문제는 도서관 간의 협력과 분담으로 해결될 수 있다.

주제어: 메타데이터, 검색엔진, 목록, 전자정보자원

ABSTRACT

This study discussed the problems of internet search engines or metadata and the necessity to catalog the digital information resources including information resources on the networks in libraries. The internet search engines or metadata make digital information resources indexed or organized immediately. But traditional library bibliographic tools like catalog should be described accurately and fully, and standardized in addition to the rapidity. Therefore from the library standpoint, digital information resources should be cataloged fully too. Cataloging cost of the increasing digital information resources can be reduced by cooperating and sharing the works among the libraries.

Key Words: Metadata, Internet Search Engines, Catalog, Digital Information Resources

---

* 이 논문은 2004년 동의대학교 연구년 지원에 의해 작성되었음
** 동의대학교 인문대학 문헌정보학전공 교수(thdoh@deu.ac.kr)
• 접수일: 2005년 2월 18일 • 최초심사일: 2005년 2월 20일 • 최종심사일: 2005년 3월 7일

Ⅰ. 시작하는 글

컴퓨터와 정보기술의 도입은 다소 고전적 이미지로 각인되어 있던 도서관과 문헌정보학계에 획기적 전환점을 가져올 수 있을 것이라는 희망으로 다가왔다. 이로 인하여 도서관 자료의 수집과 조직, 보관, 이용에 이르기까지 일상적인 도서관업무가 자동화되었을 뿐 아니라 도서관이 갖는 막연한 복고적 이미지를 첨단의 이미지로 바꿀 수 있을 것으로도 기대되었기 때문이다.

그러나 근본적으로 컴퓨터나 정보기술은 도서관이라는 개체 요소에 초점이 맞추어지거나 도서관이 주도할 수 있는 성질의 것은 아니었으며 오히려 범사회적 행보를 좌우하는 거대 기술이며 일반 현상이었다. 시간의 흐름에 따라 컴퓨터나 정보기술은 도서관의 이상이나 지향하는 목적과는 상관없이 독자적 행보를 계속해 나가고 있다. 그 결과 컴퓨터와 네트워크를 통해 속도 위주로 유통되는 정보나 정보유통 체계가 인류의 지적 유산을 집단적으로 수집, 전달, 보존하는 도서관의 목적과는 상치되거나 경쟁적 관계로 바뀔 수도 있게 되었다.

정보와 그 유통경로를 사이버 공간으로 흡수하여 빠르고 정확하게 전달하며, 쉽게 갱신하거나 폐기할 수 있도록 하는 컴퓨터와 정보기술의 끝없는 질주는 인류의 지적 창조물들을 빠르고 널리 배포될 수 있도록 하는 외에 이들의 옥석을 구분하고 체계적으로 축적, 보존하여 후세에 전달하는 도서관 본연의 사명에는 별로 관심이 없는 듯하다. 이러한 관점에서 냉정히 우리 도서관계의 현실을 돌아보면 우리가 지금 정보기술과 네트워크의 논리에 너무 깊이 빠져들고 있지는 않은지, 도서관이라는 사회적 장치와 이를 뒷받침하는 문헌정보학의 정체성은 확고한지에 대한 의구심을 가져보지 않을 수 없다.

컴퓨터와 정보기술이 도서관과 문헌정보학의 환경변화에 영향을 미친 것과 마찬가지로 도서관 자료 조직의 환경 역시 크게 바꾸어 놓았다. 책자나 카드형태를 거쳐 이제 목록이 컴퓨터 데이터 베이스 속으로 읽혀 들어가게 되었다. 이름하여 기계가독목록(MARC, Machine Readable Catalog)으로 바뀐 것이다. 한 걸음 더 나아가 정보매체의 디지털화와 이들의 양적 증대, 네트워크를 통한 정보유통의 일반화는 전자정보자원[1]을 조직하는 서지 도구로 MARC 레코드와 같은 완전한 목록 대신 가급적 간략하고 단순한 방식으로 이들 정보자원에 대한 서지 도구를 작성하기 위하여 메타데이터 스킴이나 메타데이터를 사용하는 시스템을 추구하기에 이르렀다.

전자정보자원의 서지 도구로서 시간과 비용이 많이 드는 MARC와 같은 목록을 지양하고 전자문서 작성자가 문서 자체에 포함시켜 놓은 데이터를 근거로 작성되는 서지 레코드로 대체할 수 있다고 생각하는 것이다. 그 결과 다양한 메타데이터의 표준들이 제시되고 있으며 특히 더블린코어와 같은 메타데이터 스킴에 대해서는 도서관계가 깊은 관심을 보이고 있다. 뿐만 아니라 MARC 레코드를 메타데이터로 변환하는 방안들도 연구되고 있다. 말하자면 도서관에서 목록은 그 형태의 변화

1) 이 글에서 '전자정보자원'이라 함은 디스크, 테이프 등과 같은 물리적 전자매체에 수록되었거나 네트워크를 통해 유통되는 모든 정보자원을 포괄하는 용어로 사용한다.

를 넘어 드디어 소멸을 걱정해야 할 단계에 이르게 된 것이다.

정보자원의 디지털화는 목록을 도서관 역사의 뒤 안으로 밀어내고 말 것인가? 메타데이터가 도서관 목록의 기능을 완벽하게 대체할 수는 있는 것인가?

고어먼(Michael Gorman) 이 정보기술의 발전을 감안한 신도서관학 오법칙[2] 에서 말했듯이 도서관 봉사의 질을 높이기 위해 적절한 기술을 사용함은 당연하다. 그런 의미에서 정보기술을 경원한다는 것은 천 부당하다. 오히려 적절한 활용이 권장되어야 한다. 그러나 다른 한편으로 역시 고어먼이 충고하고 있는 대로 '과거를 존중하며 미래를 창조'해야 함도 당연하다. 과거에 기반하지 않은 현재나 미래는 없다. 과거나 현재는 새로운 미래의 출현으로 말미암아 청산되어야 할 대상이라기보다 오히려 그 창조의 바탕이 될 수 있어야 하는 것이다.

목록에 컴퓨터나 정보기술을 적절히 활용함으로서 자료의 기술이나 검색, 목록레코드의 통합과 공유 등에서 그 효용성을 극대화 할 수도 있다. 그러나 도서관에서 그 역사이래 지속되어 온 목록 고유의 역할과 필요성도 여전히 계속되고 있다. 정보기술의 도입으로 말미암아 목록이 도서관의 서지 도구로서 갖는 본연의 역할은 백안시된 체 기술의 논리에만 의존한 새로운 서지 도구로 대체될 수 있다고 하는 것은 재고되어야 할 것으로 생각된다.

본 연구에서는 먼저 오늘날 전자정보자원의 서지 도구로서 빈번하게 논의되고 있는 메타데이터를 비롯한 인터넷 검색엔진의 특징을 살펴본다. 이들과 도서관의 전통적 서지 도구로서 목록이 갖는 기본 원칙과 역할들을 대비하여 봄으로써 그 특징적 차이점과 메타데이터 및 검색엔진의 도서관 정보자원의 서지 도구로서 한계성을 구명하고 이를 토대로 전자정보자원의 서지 도구로서도 여전히 완전한 목록을 작성할 필요성이 있음을 고찰한다.

그리고 원천적으로 목록 레코드를 작성하는데 수반되는 고비용의 문제에 더하여 전자정보자원의 양적 증가로 인한 전자정보자원의 목록 비용의 증대 문제를 해결하기 위하여 전자정보자원의 수집 단계에서부터 조직에 이르기까지 체계적 분담의 필요성을 살펴본다. 나아가 현재 이러한 전자정보자원의 수집과 목록에 관한 협력이나 분담이 체계적으로 이루어지지 못하고 있는 우리 나라의 실정에서 이를 실행할 수 있는 구체적인 방안을 제시해보고자 한다.

---

2) 고어먼은 랑가나단(S. R. Ranganathan)의 도서관학 오법칙(Five Laws of Librarianship) 정신을 계승한 신도서관학 오법칙(Five New Laws of Librarianship)을 다음과 같이 규정하였다.
① Libraries serve humanities
② Respect all forms by which knowledge is communicated
③ Use technology intelligently to enhance service
④ Protect free access to knowledge
⑤ Honor the past and create the future(Michael Gorman, "Five New Laws of Librarianship," *American Libraries*, Vol.26, No.8(Sep, 1995), pp.784-785.)

Ⅱ. 전자정보자원의 새로운 서지 도구들

정보기술의 발달과 병행하여 인쇄매체로 유통되던 정보자원의 상당 부분이 디지털화 되거나 네트워크를 통하여 유통 소비되고 있다. 이러한 전자정보자원들은 인쇄매체와 다른 몇 가지 특성을 갖는다. 먼저 서지적 특성들이 인쇄매체와는 다를 뿐 아니라 내용의 신빙성이 검증되지 않았거나 그 책임 소재가 명확하지 않은 경우도 있다. 또한 시사성을 띄거나, 특히 특정 시점에서 네트워크 상에 나타났다가 특별한 예고도 없이 사라져버리는 소위 '휘발성'을 갖고 있는 것들도 허다하다.

이렇게 시시각각으로 나타났다가 예고 없이 사라지거나 수시로 갱신되는 네트워크 상의 문서들을 출현 시점으로부터 최단시간에 검색하고 이용할 수 있도록 고안된 서지 도구들이 나타나고 있다. 이들을 두 유형으로 나누면 하나는 인터넷상의 정보를 키워드나 주제별로 검색할 수 있도록 색인하는 '검색엔진'이며, 다른 하나는 전자정보의 생산단계에서 해당 정보 자체에 검색을 위한 일정한 구조의 데이터들이 포함되도록 하고 이러한 데이터를 추출하여 조합함으로써 전자정보자원의 서지 레코드를 작성할 수 있도록 한 '메타데이터'이다.

검색엔진은 인터넷상의 웹사이트나 웹 페이지, 주제어 등을 검색할 때 가장 일반적으로 이용되고 있으며, 그 이름이 우리에게 낯익은 것들로는 야후(Yahoo)나 알타비스타(AltaVista), 라이코스(Lycos), 구글(Google)과 같은 것들이 있다. 이러한 검색엔진들은 스파이더나 웹 크롤러와 같은 소프트웨어를 사용하여 웹 문서상의 메타(meta) 태그를 이용하거나 빈도 등을 고려하여 키워드들을 수집한다. 이렇게 수집된 색인어들은 데이터 베이스로 만들어지며, 결과적으로 이용자들은 이 데이터 베이스를 검색함으로써 해당 웹사이트나 해당 색인어가 포함된 웹 문서를 찾아 갈 수 있게 된다.

인터넷 검색엔진은 검색방법에 따라 크게 주제어 검색엔진, 디렉토리형 검색엔진, 그리고 혼합형 검색엔진의 세 가지 유형으로 구분된다. 주제어 검색엔진은 알타비스타와 같이 인터넷상의 웹 페이지를 자동으로 색인하여 특정 키워드, 즉 주제어로 검색을 할 수 있다. 디렉토리형 검색엔진은 야후와 같이 특정 주제별로 웹 분석 전문가가 직접 웹 페이지들을 분석하여 계층별로 분류하여 둠으로써 이용자가 주제별 또는 메뉴 형식으로 검색할 수 있다. 혼합형 검색엔진은 익사이트(Excite), 라이코스(Lycos), 웹 크롤러(Web-Crawler), 인포시크(InfoSeek)와 같이 주제어와 디렉토리를 함께 사용하고 있는 것이다.[3)]

한편, 인터넷의 보편화와 더불어 도서관자료 혹은 정보의 조직을 논할 때는 어김없이 '메타데이터'(metadata)라는 용어가 등장하기 시작했다. 이 용어의 구조를 분석해보면 '~에 관한'(about)이라는 의미의 'meta'와 자료를 뜻하는 'data'가 합성된 것이다. 결국 메타데이터의 문자적 의미는 '데이터에 관한 데이터' 라는 의미가 될 것이다.

---

3) 김정현, "디지털 정보자료 관련 목록규칙의 동향," 정보관리연구, 제34권, 제3호(2003, 9), p.7.

문헌정보학과 도서관계에서는 이미 정보의 유형을 일차정보와 이차정보 등으로 구분하여 왔다. 말하자면 정보의 실체, 즉 내용(content)을 포함하고 있는 것을 일차정보라 하고 이들 일차정보에 접근하거나 식별할 수 있도록 하기 위해 만들어지는 도구적 정보, 즉 서지정보류들을 이차정보라 하였다. 이차정보란 일차정보의 식별과 접근에 긴히 사용될 수 있는 항목(데이터)들을 일정한 규칙에 따라 기술한 레코드를 리스트 한 것으로 목록이나 색인, 기타 서지류와 같은 '정보에 관한 정보'를 뜻하는 것이다. 이를 메타데이터의 문자적 개념과 연관시켜 해석해보면 도서관의 목록이나 기타 서지류들은 이미 문자적 의미에서 메타데이터의 범주에 포함되는 것이라고 말 할 수 있다.

현재 메타데이터는 문자적 의미대로 도서관의 목록과 같은 서지류들을 포괄하는 뜻으로 해석되기도 하지만 일반적으로는 이러한 의미보다는 훨씬 축소되어 네트워크를 통하여 유통되는 '전자정보자원의 서지 기술을 위한 데이터'라는 의미로 사용되고 있다. 즉 네트워크 등을 통하여 유통되는 전자 문서에 정보의 생산자가 미리 형식과 구조가 단순화된 메타데이터들을 포함시키도록 하고 이러한 메타데이터들을 기초로 전자정보자원의 서지 레코드를 신속하게 작성할 수 있도록 한 것을 뜻한다.

이러한 의미의 메타데이터를 이용하여 서지 기술을 할 수 있는 가장 보편적인 정보자원은 웹페이지를 비롯한 네트워크 상의 문서이지만 기타 물리적 매체에도 점차 확산되고 있다. 메타데이터의 유형은 기술용 메타데이터(descriptive metadata), 관리용 메타데이터(administrative metadata), 구조용 메타데이터(structural metadata) 등[4]으로 구분된다.

메타데이터 스킴(metadata scheme)이란 메타데이터 요소와 특정 목적을 위해 정의된 사용 규칙의 집합이다.[5] 메타데이터 스킴은 TEI(Text Encoding Initiative) 헤더, Dublin Core, URC 템플리트, IAFA 템플리트와 같은 것들이 있으며 그 외에도 CDWA(Categories for the Description of Works of Art), ONIX(Guidelines for Online Information Exchange) International 등 일반

4) 이들 각각을 좀 더 자세히 설명하면 다음과 같다.
- 기술용 메타데이터(descriptive metadata) : 발견(discovery, 어떤 자원을 어떻게 찾아내는가), 식별(어떤 자원을 다른 유사한 자원과 어떻게 구별할 수 있는가), 선정(어떤 자원이 특정의 요구, 예를 들면 비디오 녹화자료의 DVD 버전용이라는 요구를 충족시킨다는 사실을 어떻게 결정하는가)이라는 목적을 충족시켜주는 것. 기술용 메타데이터는 또한 병치(collocation, 어떤 저작의 모든 버전들을 함께 모으는 것)와 수집(acquisition, 특정 자원을 입수하거나 그것에 접근하는 것)을 위해서도 사용할 수 있다. 여기에 포함되는 그 밖의 기능들로는 평가(evaluation) 링크(linkage) 가용성(usability) 등이 있다.
- 관리용 메타데이터(administrative metadata) : 자원의 관리를 용이하게 하기 위한 정보이다. 이 메타데이터는 어떤 객체가 언제 어떻게 작성되었고, 컨텐트(내용)에 대한 접근을 관리하거나 컨텐트를 아카이빙(arcniving)하는 책임을 누가 가지며, 그와 관련하여 어떤 통제나 처리가 수행되고 있고, 접근이나 이용에 대한 어떤 제약이 적용되고 있는지와 같은 정보를 포함한다.
- 구조용 메타데이터(structural metadata) : 복합적인 디지털 객체들을 함께 묶어주는 접착제로 생각할 수 있다. 예를 들면 책은 여러 장으로 이루어지고, 각 장은 여러 페이지로 이루어지며, 각 페이지는 별도의 디지털 파일로 나타날 수도 있을 것이다. 구조용 메타데이터는 물리적인 파일과 페이지, 페이지와 장, 장과 전체로서의 책 사이의 관계를 기록하기 위해 필요하다.(Priscilla Caplan 저, 메타데이터의 이해, 오동근 역(대구 : 태일사, 2004), pp.25-28.)

5) 상게서, p. 29.

에게는 비교적 생소한 것들에 이르기까지 그 목적과 용도에 따라 다양한 종류들이 나타나고 있다.

그 중에서도 문헌정보학 분야에서 가장 주목받고 있는 메타데이터 스킴은 더블린 코어(Dublin Core)라고 할 수 있다. 더블린 코어는 네트워크 상의 정보자원을 기술하는데 필수적인 데이터 요소를 도출하기 위해 1995년 OCLC와 NCSA에 의해 더블린에서 개최된 워크샵에서 채택된 것이었으므로 이러한 이름을 갖게 되었다.

Ⅲ. 네트워크 세계와 메타데이터의 빛과 그늘

인터넷 보급이 확산되고 전자정보자원의 유통이 증가되면서 도서관에서 전자정보자원의 서지도구로써 메타데이터에 거는 기대가 점차 높아지고 있다. MARC와 같은 도서관의 전통적 서지도구가 메타데이터로 대체될 수 있을 것으로 주장되기도 한다. 여기서는 전자정보자원과 메타데이터에 대한 평가의 허와 실을 검토해보고자 한다.

1. 전자정보자원의 특성과 도서관의 역할

인터넷으로 대표되는 네트워크의 세계를 흔히들 정보의 보고라고 극찬하거나 이를 굳게 믿고 있는 사람들이 많다. 인터넷상에서 사이트를 옮겨다니며 항해를 하다보면 그 끝을 찾을 수 없는 정보의 바다에 접할 수도 있게 된다. 급격하게 늘어나고 있는 정보자원들을 관리해야 하는 부담은 작고 가볍고 빠르게 축적, 전달될 수 있는 전자 매체와 사이버 공간을 통하여 해결될 수 있을 것이라고 말하기도 한다. 그래서 많은 사람들이 인터넷을 만능의 해결사라고 생각할 수도 있을 것이며 유용한 도구임도 부정할 수는 없다.

그러나 조금 더 유심히 인터넷의 이면을 관찰해보면 이러한 사이버 공간상의 정보들이 갖는 문제점들도 쉽게 발견할 수 있다. 혹평하는 사람들의 표현을 빌리자면 '정보의 쓰레기 더미'라고 표현될 정도의 온갖 잡다한 것들, 정보라는 용어를 붙여야 할지조차 의심스런 잡동사니들도 무질서하게 남발되고 있음을 볼 수 있다.

어느 정도 외적인 형식은 갖추었다고 하더라도 아직은 그 신빙성과 역사성이 증명되지 않은 문서들도 많다. 사이버 공간에서 정보를 생산하고 배포하는 데는 특별한 심사과정이 없는 경우가 대부분이며 정보를 생산하거나 배포하는 과정에서 최소한의 거름장치도 없는 경우가 허다하다. 이렇게 생산 배포된 정보들은 근본적으로 신빙성의 문제뿐만 아니라 무가치한 것 혹은 악의적인 것조차 있을 수 있다.

또한 네트워크 상의 정보는 "작고 가볍고 잘 섞인다"[6]는 표현으로 묘사될 만큼 수정과 갱신, 삭제가 자유롭다. 이들 정보는 수시로 바뀔 수도 있으며 어느 날 갑자기 네트워크 상에서 사라져

버릴 수도 있다. 갱신된 이후에는 이전의 자료가 잘 보존되지 않는다. 인쇄 매체에서와 같이 시간의 흐름에 따른 내용상의 변화를 확인할 수 있거나 역사적으로 보존되는 경우는 흔치 않다. 사이버 공간에서의 정보생산과 유통이 점증하는 시대적 추세 속에서 인터넷 환상 속에 매몰된 이러한 문제들을 간과한다면 가까운 미래에 인류는 그 역사를 상실한 채 허공을 딛고 불안한 현재와 미래를 맞이하게 될 수도 있을 것이다.

결국 이러한 사이버 공간상의 정보는 어떤 방식으로든 '알곡'과 '가라지'로 선별되는 과정을 겪어야 할 것이다. 정보로서의 가치나 지적 문화적 유산으로 보존되어야 할 가치가 있는지를 검증 받아야 하며 이는 실물 정보자원의 경우에도 평가와 선택의 과정을 거쳤던 것과 같은 맥락이기도 하다. 검증을 거친 것들에 대해서는 장기적 이용과 보존의 수단이 강구되어야 한다. 통제되기 어려운 소위 정보의 쓰레기더미로부터 옥석을 구분하여 가치 있는 정보자원들을 추출하고 이를 이용자 일반에게 제공하거나 장기적으로 보존할 수 있는 체제가 마련되어야 한다.

일찍부터 고어먼은 인류의 기록을 보존하여 이를 다음 세대에 전승하는 역할은 도서관 이외에 맡을 기관이 없다는 사실을 다음과 같이 적시한 바 있다.

> 도서관의 가장 중요한 임무는 미래의 세대를 위하여 인류의 기록을 보존하는 일일 것이다. 이것은 엄숙한 임무이다. 왜냐하면 미래의 인간들은 우리가 보존하여 전해 주는 것만을 알 수 있을 것이기 때문이다. 더구나 도서관 이외에는 오로지 기록된 지식과 정보를 보존하는 데 매달리고 있는 사회조직이란 따로 없다. 서점, 출판사, 박물관, 정부, 컴퓨터회사, 그 밖의 어떤 사회조직도 도서관과는 이점에서 다르다. 인류의 기록을 보존하는 데 도서관은 아주 훌륭한 기록을 세우고 있다.[7]

미래의 세대를 위하여 인류의 기록을 축적하고 보존하는 일은 비단 인쇄매체를 비롯한 실물장서에 국한되는 것은 아닐 것이다. 정보기술과 컴퓨터로 대변되는 시대적 상황은 실물자료의 상당부분이 전자정보자원으로 대체될 것임을 시사하고 있으며 이와 같은 전자정보자원 역시 '미래 세대를 위한 인류의 기록'의 일부임에 틀림없기 때문이다.

따라서 도서관은 실물장서와 마찬가지로 전자정보자원의 수집과 보존에도 관심을 기울여야 할 때가 된 것이다. 도서관이 수집하고 보존해야 하는 '미래 세대를 위한 인류의 기록'의 범주에는 전자정보자원도 포함되어야 하며 '보존되어야 할 정보자원'으로서의 가치 검증도 역시 도서관의 몫임에 틀림없을 것이다. 이렇게 함으로써 정보사회에서도 도서관이 인류의 지식정보자원들을 수집 축적 보존하고 이를 후세에 전달하는 책임 있는 사회적 장치로서의 소임을 다 할 수 있게 될 것이다. 작금의 추세로 미루어볼 때 이들 전자정보자원에 대한 수집과 보존 대책은 서두르지 않으면 늦게

6) 우병현, 디지털은 자본이다 : 디지털자본시대의 정보민족주의(서울 : 나남출판, 1996), p.67.
7) 마이클 고어먼 著, "21세기의 도서관과 사서직 : 태평양 연안국들을 중심으로," 김정근譯, 도서관, 제52권, 제3호(1997, 가을), p.129.

될 것이며, 특히 쉽게 사라져 버릴 수 있는 전자정보자원의 특성상 이러한 대책은 계획적이며 체계적으로 수립되고 실행되어야 한다.

2. 검색엔진과 메타데이터의 문제점

인터넷 검색엔진이나 메타데이터가 전자정보자원을 적시에 신속하게 검색할 수 있도록 하는 도구로는 요긴하게 사용될 수 있을 것이다. 그러나 이들 검색엔진이나 메타데이터가 도서관 정보자원의 전반적 관리를 위한 서지 도구로서는 몇 가지 문제를 가질 수 있다.

각각의 검색엔진들이 인터넷이라는 거대한 정보유통의 장을 시간과 공간적으로 망라하기는 어려울 것이다. 네트워크 상의 정보가 갖는 가변성과 휘발성 때문에 계속해서 원 정보를 추적하고 그에 따라 지속적으로 갱신되지 않으면 이용자들로 하여금 낭패를 겪게 하는 경우가 다반사이다. 어떤 검색엔진에 의해 추출된 웹사이트나 문서를 추적해보면 이미 그 사이트가 폐쇄되어 찾을 수 없거나 사이트가 존재하더라도 해당 문서가 이미 사라져버리고 없는 경우를 흔히 접하게 되는 것이 그 예이다.

또한 질적인 면에서 효율적이거나 완벽한 검색도구가 되는 것도 쉽지는 않을 것이다. 검색엔진은 근본적으로 정보생산자에 의해 작성된 키워드나 데이터에 의존하여 색인하는 경우가 대부분이다. 그러므로 어휘 통제에 문제가 있을 수 있으며, 뿐만 아니라 문서 작성자의 고의에 의해 불필요하거나 악의적 정보가 검색될 수 있는 색인어가 추출되거나, 또는 색인이 불가능하게 되는 웹 페이지들도 있을 수 있다.

한편 村上泰子는 전자정보자원의 메타데이터가 그 특성상 목록에 미치지 못하는 문제점에 대하여 다음과 같은 세 가지로 지적하고 있다.

> 첫째, 메타데이터는 구문(syntax) 부분과 의미(semantic)가 완전히 별개로 규정되어 있으며, 둘째, 메타데이터에는 현재 각 요소에 기술해야 할 내용 등에 있어 메타데이터의 종류에 관계없이 적용할 수 있는 공통입력 매뉴얼과 같은 것이 존재하지 않고, 셋째 메타데이터의 작성이 정보소비자 측이 아니라 생산자 측에서 이루어지고 있다는 것을 전제로 하고 있는 점이다.[8)]

특히 전자문서의 작성단계나 메타데이터의 서지 레코드를 만드는 단계에서 준용해야 할 메타데이터 스킴들이 그 목적이나 용도에 따라 매우 다양하다. 말하자면 전통적으로 인쇄매체의 목록을 작성하기 위한 서지기술규칙이나 MARC 형식과 같이 모든 유형의 전자문서에 공통적으로 적용될

8) 村上泰子, "圖書館界とメタデ-タ : 米國議會圖書館の戰略を中心に," 情報の科學と技術, Vol.51, No.8(2001), pp.402-408, 김정현, 전게논문, p.9에서 재인용.

수 있는 매뉴얼이 없다는 것이다. 그나마 도서관계에서 큰 관심을 갖고 있는 메타데이터로는 더블린 코어 정도를 들 수 있는데 이 역시 모든 전자문서의 작성자들이 이 것에 의해 메타데이터를 삽입(embed)한다고 기대할 수는 없다.

더블린 코어는 그 특성이 1) 데이터의 형식과 구조를 단순화하여 원문의 저자나 발행자가 메타데이터를 직접 작성하고, 2) 네트워크 출판을 위한 저작도구의 개발자가 이 정보에 대한 템플릿을 직접 소프트웨어에 포함할 수 있도록 하며, 3) 작성된 데이터를 기초로 특정 분야에서 요구되는 상세한 수준으로 확장하여 사용하기 위한 것[9]이라고 요약되고 있다.

이와 같이 처음부터 메타데이터는 네트워크 자원의 서지 레코드를 간단하고 기계적으로 작성하는 것을 목적으로, 정보생산자가 직접 쉽게 기술할 수 있도록 서지 데이터의 형식과 구조를 단순화한 것이었다. 결국 전자정보자원의 메타데이터나 이것에 의한 서지 레코드의 작성은 천차만별한 배경의 저자들로 하여금 책을 저술하거나 발행하는 단계에서 미리 간략하게 서지 데이터를 작성하도록 권장하고 이것들을 조합하여 목록을 작성하도록 한 것에 크게 다르지 않게 되는 것이다.

이는 목록과 같은 도서관의 서지 도구를 너무 단순하게 이해한 결과가 아닌가 생각된다. KCR이나 AACR과 같은 전통적 목록규칙의 표목과 기술규칙에서 자료의 유형과 기술 데이터의 종류별로 일일이 규정된 복잡한 규칙과 원칙들, 그리고 이러한 규칙들에 부가되고 있는 해설들을 보라. 사서와 같은 정보조직의 전문가가 아닌 정보 생산자에 의해 작성되고 형식과 구조가 단순화된 메타데이터가 목록규칙이나 MARC 형식과 같은 정밀하고 복잡한 규정들에 의해 작성되는 도서관의 서지 레코드가 갖추어야 할 요건들을 모두 만족시킬 수 있을까? 이것들이 '완전성'과 '정확성', 그리고 '표준화'가 아무리 강조해도 지나치지 아니하는 도서관의 서지 데이터로서 적합할 수 있을까? 그리고 이러한 메타데이터에 의한 서지 레코드가 실물장서의 MARC 레코드와 체계적으로 통합되고 도서관간에 공유될 수는 있을까?

뿐만 아니라 메타데이터는 인터넷 검색엔진에서와 마찬가지로 정보생산자에 의해 고의로 조작될 가능성도 배제할 수 없다. 정보의 생산자에 의해 메타데이터가 의도적으로 조작될 가능성에 대해 캐플런(Priscilla Caplan)은 다음과 같이 지적하고 있다.

> 스패밍(spamming)은 서치 엔진의 검색에 영향을 주기 위해 웹 페이지에 키워드를 지나치게 많이 부여하는 것이다. 이와 같은 키워드들이 이용자에게 보이지 않도록 하기 위해, 그러한 키워드들은 종종 페이지의 배경색과 같은 색으로 입력하거나 〈meta〉태그들의 컨텐트로서 입력한다. 예를 들면 자동차 부품 도매업자는 자신의 홈페이지의 등급을 높이기 위해 "automotive"와 같은 단어들을 수 백 번 반복할 수 있을 것이다. 더 나쁜 경우로는, 포르노 사이트가 자동차 부품을 탐색하는 사람들의 결과 세트에 포함되도록 하기 위해 그 키워드 사이에 "automotive supplies"와 같은 용어들을 포함시킬 수도 있다는 것이다.[10]

9) 김태수, 더블린코어(Dublin Core), http://dewey.yonsei.ac.kr/metadata/dc.htm [인용 2004. 7. 31]
10) Priscilla Caplan 저, 앞의 책, pp.101-102.

네트워크 상의 전자문서를 생산하는 주체가 문서의 내용과 관계없이 이용자들이 빈번하게 사용하는 키워드들을 의도적으로 사용함으로써 실제로 검색하는 내용과는 무관한 문서나 웹사이트가 검색 대상에 포함되게 하거나 혹은 그와 반대로 검색대상에서 제외되도록 조작할 수도 있다는 것이다. 이는 결과적으로 검색의 효율을 떨어뜨릴 뿐 아니라 경우에 따라서는 악의적 정보를 광범하게 배포되게 하는 결과를 가져오게 하기도 한다.

이와 같이 현실적으로 메타데이터가 정보생산자의 의도에 따라 조작될 수 있으며, 또한 조작되는 경우도 있다면 전적으로 메타데이터에 의존하여 서지 레코드를 작성하는 것은 그 신뢰성에 문제가 있을 수 있을 뿐 아니라 자칫 정보생산자의 악의적 함정에 빠질 수도 있게 되는 것이다. 결과적으로 목록과 같은 도서관의 서지 레코드가 갖는 기능과 역할을 감안할 때 이를 메타데이터로 대체하는 것이 가능한지에 대하여 의심을 갖지 않을 수 없다.

Ⅳ. 도서관 목록에 대한 재인식

전자정보자원의 검색이나 조직을 위한 노력들은 다음과 같이 구분될 수 있다. 인터넷상의 정보검색을 위한 검색엔진 외에 전자 문서에 포함된 메타데이터를 추출하고 이를 조합하여 서지 레코드를 작성하려는 시도, 그리고 목록규칙이나 MARC와 같이 전통적인 도서관의 자료조직 도구들을 전자정보자원에도 적용할 수 있도록 개정하려는 시도들이 그 것이다.

특히 근래에 각종 전자정보자원을 기술하는데 기존의 목록규칙이나 MARC 형식을 사용하는 것은 그 자체가 갖는 구조적인 문제나 규칙의 복잡성 때문에 적절하지 못하다는 논의들이 대두되고 있다. MARC 형식을 사용하여 서지 레코드를 작성하는데는 시간과 비용이 많이 소요되기 때문에 메타데이터를 사용하여 서지 레코드를 작성하는 것이 유리하다는 것이다. MARC 레코드 작성에 소요되는 고비용 문제는 MARC가 전자정보자원의 기술에 적용되기 이전부터 지적되어 온 것으로 전자정보자원의 증가 추세를 감안할 때 이를 수집 관리해야 하는 도서관의 목록 비용은 더욱 증가하게 될 것이다. 이러한 전자정보자원의 목록 비용과 시간의 문제를 메타데이터가 해결해 줄 수 있을 것으로 기대하고 있는 것이다.

티커슨(David A. Tyckoson) 같은 사람은 전통적인 도서관의 목록을 새로운 운송 수단들의 등장으로 승객의 대부분을 잃게 된 호화열차 "The Twentieth Century Limited"[11]에 비유하여 이용자들이 목록 이외의 다른 경로를 통하여 보다 많은 정보를 발견할 수 있도록 해주는 '기술 발달'의 희생자가 될 것이라고 했으며, 또한 목록의 구조를 18세기의 이상을 실현하기 위하여 21세기의 기술을 사용하는 형상이라고 까지 하였다.[12]

11) 뉴욕과 시카고를 연결하는 가장 빠르고 호화로웠던 열차로 당시로는 다른 어떤 교통수단보다도 선호되었던 교통수단이었음.

말하자면 정보 매체의 다양화, 특히 네트워크를 통한 전자정보자원이 증가해 가는 시대적 상황에서 이들의 서지 기술을 위해서는 새로운 방식의 도구들이 필요하며, 목록과 같은 전통적 서지 도구의 획기적 변화가 요구됨을 시사한 것이다. 이는 네트워크의 보급 확대 와 더불어 전통적 도서관 목록의 종언, 혹은 최소한 전자정보자원에 대해서는 목록이 더 이상 불필요하다는 의미로 해석될 수도 있다.

이러한 관점에서 메타데이터는 기존의 MARC나 목록규칙과 경쟁적 관계를 갖게 될 것이다. 이미 전자정보자원에 대하여는 메타데이터에 의한 서지 도구의 작성이 불가피하다고 주장되거나, MARC 데이터를 메타데이터로 변환하거나 메타데이터로 통합하기 위한 연구와 실험들도 이루어지고 있다.

더블린코어가 웹 자원의 주된 메타데이터 기술 수단으로 사용되고 있는 이유를 1) 레코드 구조의 단순성 2) 색인의 단순성 3) 이용의 용이성 4) 상호 운용성 5) 자원의 접근성[13] 등으로 들고 있다. 메타데이터는 이와 같이 신속하고 간단하게 서지 레코드를 작성할 수 있게 함으로써 정보자원의 신속한 검색도구나 검색엔진으로서의 역할은 무난히 소화할 수 있을지도 모른다. 그러나 목록이 갖는 고유한 역할과 기능이 여기에 그칠 수는 없다. 도서관의 서지 레코드는 단순하고 신속한 검색도구로서의 기능 외에도 고유의 다양한 기능들을 갖고 있다. 이와 같은 목록의 기능과 역할은 전자정보자원을 위한 서지 도구에서도 여전히 간과되어서는 안 될 것들이다.

근본적으로 도서관의 목록이 지향해야 할 기능과 역할에 대하여 현규섭은 다음과 같은 원칙으로 강조한 바 있다.

> 제1원칙(역사성의 원칙) : 서지기술은 한 문헌이 역사적으로 존재하였음을 기록하는 사실증명 행위이다. 기술된 목록은 한 시대, 한 시점에서 발생한 문헌의 존재를 기록하는 행위의 소산이어야 한다. 이 기록은 영구적인 것으로서 설혹 문헌 그 자체가 인멸되더라도 기술된 목록만은 남는다는 사실에 기초되어야 한다. …
>
> 제2원칙(사실통지의 원칙) : 서지기술은 문헌에 나타나 있는 사실을 있는 그대로 기록하여 이를 통지하는 작업이다. 문헌에 나타나 있는 사실이 다음 세대에 또는 타인에게 있는 그대로 통지되어야 한다. …
>
> 제 3원칙(소거가치의 원칙) : 기술되는 요소의 배치순서는 한 요소가 소거될 때 미치는 전체 기록에 대한 가치(즉 소거의 효용가치)로서 판정한다. …
>
> 제4원칙(식별성의 원칙) : 소거가치가 동등한 요소의 배치순차는 식별의 효과로서 차등지운다. …
>
> 제5원칙(표준성의 원칙) : 한 목록자에 의하여 작성된 서지기술은 다른 목록자에 의하여 재 반복되더라도 언제나 동일한 결과가 나타나야 한다.[14]

---

12) David A. Tyckoson, "The Twenty-First Century Limited : Desighing Catalogs for the Next Century," *Cataloging and Classification Quarterly*, Vol.13, No.3(1991), p.27.

13) Hanses, Preben, User guidelines for Dublin core creation, 1998, 〈http:www.sics.se/ ~preben/DC/DC_guide.html〉, 金泰樹, "MARC 데이터의 메타데이터 변환연구," 國會圖書館報, 第35卷, 第6號(1998. 11, 12), p.4에서 재인용.

14) 玄圭燮, "書誌記述의 原則確立을 위한 覺書," 도서관, 제38권, 제1호(1983), pp.6-7.

목록에 의한 정보자원의 서지기술은 일차정보원의 손·망실에도 불구하고 그 존재 사실을 증명할 수 있을 만큼 자세하여야 하며, 또한 기술사항의 중요성에 따라 결정된 순서와 통일된 형식으로 기술되어야 함을 뜻한다. 결과적으로 목록을 대표로 하는 도서관의 서지 레코드는 원 문헌의 체계적 압축임과 동시에 그 대용 가치를 갖게 되며, 또한 논리적 체계에 따라 통합되거나 재조직 될 수 있도록 형태적으로 통일되어야 하고, 동일한 문헌에 대해서는 어떤 목록자에 의해 만들어진 것이라도 동일할 수 있도록 표준화되어야 한다는 것이다.

이와 같은 도서관의 목록은 서지기술에 관한 한 비전문가이며 다양한 배경과 언어를 가진 정보생산자에 의해 산발적으로 작성된, 그리고 구조나 형식에서 단순화를 지향하는 메타데이터에 의존해서는 본래의 목적을 달성하기가 어려울 것이다.

한편 김남석은 목록의 기능을 검색기능 식별기능 소재지시기능 사전적 기능 등 네 가지로 들고 특히 목록의 사전적 기능에 대하여는 다음과 같이 설명하였다.

> 목록은 기입되는 사항을 통해서 이용자가 자료에 직접 접하지 않아도 그 자료의 내용, 형식, 발행, 형태적인 제반 사항에 대한 정보를 제공하는 한 자료에 대한 사전적인 역할을 해야 한다. 목록에 채기되는 사항에는 분류번호와 주제명이 있어 자료가 담고 있는 내용과 형식에 대한 정보를 제공하고, 서명과 저자사항을 통해 자료의 성격에 따른 각종 표제와 저작자에 대한 정보를 제공하며 판차사항을 통해 판의 개정에 관한 정보, 발행사항을 통해 출판, 인쇄 배포에 관한 정보, 형태기술사항을 통해 자료의 서지적인 정보, 그리고 주기를 통해 자료의 특징적인 영역과 내용, 요약 등의 정보까지 제공하고 있다. 이러한 의미에서 목록은 자료에 대한 사전적인 역할을 감당한다.[15]

목록의 작성에 필요한 기술사항은 이와 같이 포괄적이며 종합적인 것들이며, 또한 문헌에 따라 그 적용 여부가 달라지거나 그 형태가 일일이 예시하기 불가능할 만큼 다양하기도 하다. 도서관의 목록이나 목록규칙이 복잡하고 까다로울 수밖에 없는 것도 이에 연유하는 것이다. 이러한 상황에서 전적으로 정보생산자에 의해 만들어진 서지 데이터에 근거하여 서지 레코드를 작성한다는 것은 원천적으로 그 부실을 전제하지 않을 수 없는 것이다.

목록의 원칙이나 본질적 기능들을 종합적으로 고려할 때 목록은 단지 그 원 문헌의 위치를 추적하거나 탐색하는 도구에 그치는 것은 아니며 문헌에 대한 완전하고 정확한, 그리고 통일된 서지기술이어야 하는 것이다.

더욱이 MARC 개발 이후 MARC 데이터 베이스를 중심으로 수서, 목록, 대출 등의 하부시스템을 개발하는 도서관자동화시스템의 개발 패턴이 일반화되었다. 목록 데이터 베이스는 단순한 서지 레코드의 집적체가 아니라 도서관의 전반적인 시스템을 관리하기 위한 중추적 파일이 되어야 하는

15) 金南碩, 資料目錄學, 第3增補版(大邱 : 啓明大學校出版部, 2000), p.20.

것이다. 결과적으로 도서관의 정보자원에 대한 서지 레코드는 단순성보다는 완전성과 정확성, 표준성 그리고 도서관의 다른 서비스 및 관리시스템과의 연계성까지 고려되어야 하는 복합적인 성질을 띠어야 하는 것이다.

기존의 도서관 자료와 마찬가지로 전자정보자원 역시 정보의 가치에 따라 장기적으로 이용될 수 있도록 가공 및 축적될 필요가 있으며 전자정보자원이 네트워크 상에서 추출되어 도서관의 정보자원으로서 장기적으로 축적되고 이용되는 단계에서는 이들의 서지 레코드 역시 도서관 목록이 갖는 본연의 역할을 감당하기 위해 필요한 일체의 속성과 요소들을 구비해야 하는 것은 당연하다.

전자정보자원의 검색도구들과 목록의 관계에 대해서 고어먼은 일찍부터 다음과 같이 주장한 바 있다.

> 전자문헌과 인쇄와 같은 선형문헌간의 큰 차이는 전자문헌은 많은 경우 물리적으로 도서관 내에 문헌 자체가 없으며 성질상 변하기 쉬우며 실제로 자주 변한다는 것이다. 이것을 전제로 했을 때 분명해지는 것은 사서들은 이용자의 필요와 관심, 서비스 비용 등을 감안하여 접근을 제공하고자 하는 자료(CD-ROM, 온라인 등)에 한하여 선정할 것이라는 점이다. 역시 분명해지는 것은 사서들의 통상적인 서지통정기법(분류, 목록, 색인 등)이 상용 전자서비스에서 보이는 아마추어적인 '탐색도구'와 기타 검색기법에 비해 월등하게 우수하다는 점이다. 도서관 기법의 강점은 복잡하지만 질서정연하며 이용자 친화적인 서지통정 구조와 인터넷이나 월드 와이드 웹의 혼란한 상태를 비교해 보면 금방 알 수 있다. 문제는 어떻게 전자문헌을 조직할지가 아니라, 어떤 문헌을 조직해야 하는가 이다. 웹을 탐색해 보면 바로 알 수 있는 사실은 거기서 얻을 수 있는 것은 대부분 단명하고 별 가치 없는 내용이라는 것이다. 나는 지금 서지통정의 대상이 되는 전자문헌을 어떻게 식별할 것인가 하는 문제에 대하여 해결책을 정확히 댈 수는 없다. 하지만 사서들은 언제나 자료를 선정하고 있으며 구입여부에 대한 판단을 내리고 있다는 점을 지적해 두고 싶다. 우리가 필요로 하는 것은 혼란을 질서로 바꾸고 전자문헌을 검색하고 이용하는 것이 책, 저널, 그 밖의 도서관 자료의 경우처럼 쉽게 될 수 있도록 하기 위하여 모든 수준(개체 도서관, 지역, 국가, 국가간)에서 다양한 노력을 기울이는 것이다.[16]

또한 목록이 전자정보 시대에도 사이버 공간의 혼돈 상태를 정리할 수 있는 방법과 역할에 대해 영미목록규칙(AACR) 개정을 위한 협동조정위원회(Joint Steering Committee)의 의장이었던 허쓰웨이트(Ann Huthwaite)는 다음과 같이 단정하고 있다.

> 인터넷 정보자원들을 어떻게 목록 할 것인지는 더 이상 논란거리가 될 수 없다. AACR 2 내의 규칙들은 완벽하고 충분하다. 그리고 메타데이터 스킴들은 접근을 위한 수단으로 사용될 수 있을 것이다. 문제는 무엇을 목록할 것인가 하는 것이다. 이는 장서구성에 관계되는 것이지 목록에 관련되는 문제는 아니다. 도서관 장서개발 정책에는 지속적 가치를 가지는 인터넷 정보자원을 결정하기 위한 기준을 포함시켜야 할 것이다.

16) 마이클 고어먼 著, 전게논문, pp.127-128.

그래서 도서관의 정보자원으로 결정된 인터넷 정보자원들은 목록에 포함되어야 한다.[17]

이러한 견해들을 종합해보면 인터넷상의 정보자원을 통정하기 위한 서지 도구로서 검색엔진이나 메타데이터와 도서관의 목록이 지향하는 목표가 다름을 알 수 있다. 나타났다가 사라지기를 거듭하는 네트워크 상의 정보를 신속하게 색인하고 적시에 이용할 수 있도록 하는 정보검색 도구로서 검색엔진이나 메타데이터의 유용성은 부정할 수 없다. 그러나 이러한 검색엔진이나 메타데이터가 지향하는 역할만으로는 도서관의 목록이 갖는 고유한 기능을 완전히 수행할 수는 없을 것이다.

오늘날의 기록 문화 환경에서 도서관은 이러한 전자정보자원의 이용과 보존, 그리고 서지통정에도 관심을 기울이지 않을 수 없다. 왜냐하면 도서관은 그 사회의 지식정보자원들을 관리하고 보존하는 사회적 장치로서의 사명을 감당해야 하며, 이러한 지식정보자원의 범주에는 전자정보자원도 포함되는 것이 틀림없기 때문이다.

따라서 도서관은 네트워크 상의 정보들을 일시적으로 유통되다가 사라져버려도 좋을 것들로부터 선별한 정보자원들에 대하여 장기적 보존 대책을 강구해야 하며 또한 체계적이며 완전한 서지통정이 가능하게 해야 한다. 도서관은 이들 정보자원들에 대하여도 정확성과 완전성 그리고 표준성에 바탕한 '복잡하지만 질서정연한' 서지 도구를 유지할 필요가 있다. 이와 같은 서지 도구로서 구조와 기술의 단순화, 서지 기술 작업의 간략화를 근간으로 하는 메타데이터는 근본적으로 한계를 가질 수 밖에 없으며, 목록은 여전히 가장 이상적인 서지 도구에 틀림없을 것이다.

Ⅴ. 전자정보자원 목록의 나아갈 방향

도서관계는 정보기술의 도입과 더불어 그 고유의 핵심 영역이라 할 것조차 노동집약적이며 비용이 많이 든다는 이유로 관심 영역 밖으로 밀어내고 싶어하는지도 모르겠다. 실물장서를 가벼이 여기는 것이나 분류나 목록과 같은 도서관의 중심업무가 '아웃소싱' 등의 핑계로 우리 관심의 주변부로 밀려나고 있는 사조가 그렇다. 최성진의 다음과 같은 주장은 일찍이 이러한 현상에 대한 도서관계의 자성을 촉구한 것이었다.

분류 편목은 수 천년동안 있어온 것이며 앞으로도 오랫동안 도서관 안에 있을 것이다. 그것이 최근에 와서 갑자기 쓸모 없어지고 진부해진 일이 없다. 현재의 목록 형식이 현대인에게 맞지 않거나 분류업무를 더 효율화시킬 기술과 도구가 출현하였다면 그 형식과 방법을 개선하기 위하여 오히려 더 연구할 필요가 있을 뿐이다. 그렇게 해서 조속히 시대에 맞는 새로운 형식의 도서관봉사가 제공될 수 있게 하여야 한다. 아무리 마음에

17) Ann Huthwaite, AACR2 and Its Place in the Digital World : Near-Term Solutions and Long-Term Direction, 〈http://lcweb.loc.gov/catdir/bibcontrol/huthwaite_paper.html〉 [cited 2004. 4. 22]

> 들지 않더라도 수집, 조직, 축적, 봉사, 관리를 문헌정보학의 울타리 밖으로 던져버릴 수는 없다. 실상은 가장 아끼고 소중히 여겨야 할 것들이다. 그것들이 없어진 문헌정보학은 이미 문헌정보학이 아니다. 또 그것들만 있으면 그 새롭다고 하는 주변 주제의 연구성과가 없더라도 문헌정보학의 입지에는 흔들림이 없을 것이다.[18]

심우준도 학문적 전통을 경시하고 오로지 새로운 기술과 변화하는 환경에의 적응에만 몰두하는 작금의 상황에 대해 다음과 같이 경고한 바 있다.

> 문헌정보학 분야의 연구에서 개설류나 통계적 방법에 의한 이론의 구축 내지 기계화 처리에 주안을 두면 도서관의 기능은 마비되지 않는다고 생각하거나, 특히 목록 분야에서 고도의 문명에서 산출되는 홍수와 같은 자료를 어떻게 통정할 것인가 하는 데만 신경을 씀으로서 학문 자체가 최근에 생겨난 것처럼 착각하고 과거를 단절해 버린다면 이것은 비극이 아닐 수 없다.[19]

이러한 우려들은 오늘날의 문헌정보학계나 현장에서 도서관이 가진 고유한 사명을 유보한 채, 외형이 화려한 첨단 기술에 현혹되어 컴퓨터나 정보기술에 몰입하고 있는 사태를 예견하고 미리 경고하고 있는 듯 하다. 컴퓨터와 정보기술의 도입에도 불구하고 문헌정보학의 핵심 영역에 대한 관심과 연구의 심화를 주문하고 있으며 특별히 분류나 목록과 같은 전통적 영역들에 대해 도서관적 논리와 역할이 여전히 존중되고 추구되어야 함을 강조하고있다.

물론 도서관이라고 지난 수 백년 간 계속해 온 일들을 관습적으로 반복하거나 먼지 쌓인 서가에 갇혀 소수의 선택된 독자만을 기다려야 한다는 것은 아니다. 정보기술을 도입하고 그 혜택을 누려야 하는 것은 지당한 일이다. 다만 이러한 정보기술이나 이론의 도입은 도서관의 본질적 기능을 보다 효과적으로 수행할 수 있도록 하는 도구적 개념이어야 한다는 것이다.

우리는 이러한 도구적 개념의 변화를 본질의 변화 즉 '변질'로 착각하는 듯한 느낌을 갖게 한다. 도서관이 추구해야 할 이상과 역할은 이러한 도구적 개념의 변화에 의해 훼손될 수 없는 것이다. 이러한 맥락에서 오늘날 도서관계의 목록과 메타데이터에 대한 관심을 반추해보면 우리는 너무 쉽게 목록을 뒤로하고 빠르고 쉽고 간편한 서지 도구의 작성에 관심을 집중하고 있는 듯한 느낌을 지울 수 없다.

도서관이 그 기능을 제대로 수행할 수 있기 위해서는 목록이 갖는 본질적 역할의 계승은 불가피하다. 전자정보자원에 대한 서지 도구 역시 '간단하고' '빠르게'보다는 '완전'하고 '정확'하게 그리고 '표준화'를 지향함으로써 체계적이며 질서 정연한 서지통정이 가능하도록 해야 할 것이다.

목록에 의한 전자정보자원의 서지통정을 가능하게 하기 위해서는 무엇보다 먼저 목록규칙을 비

---

18) 崔成眞, "한국문헌정보학에 대하여," 蒼史李春熙敎授定年記念論叢(서울 : 記念論叢刊行委員會, 1993), p.91.
19) 沈嵎俊, "2000年前의 編目者들의 役割 : 우리는 무턱대고 過去를 버려야 하나?," 國會圖書館報 第17卷, 第2號 (1980, 3, 4), p.78.

롯한 MARC 형식이 전자정보자원을 수용할 수 있고 장점을 가장 잘 발휘하도록 조정되어야 할 것이다. 전자정보자원이 보편화되는 시대에도 목록이 제자리를 확보하기 위해 필요한 방안과 방향에 대하여 델시(Tom Delsey)는 다음과 같이 말하고 있다.

> 도서관 목록이 이러한 (전자정보자원이 보편화되는) 환경에서도 가장 우선적인 정보자원에의 접근도구로서 위치를 확보하기 위해서는 네트웍을 통해 광범위하게 전개되는 기술에 뿐만 아니라 목록이 다른 검색도구들과는 차별되게 갖는 역할과 그 구조의 완전함에 전략적으로 초점을 맞출 필요가 있다. 그러한 요소 가운데 하나는 목록의 과정이다. 접근 장치로서 목록의 가치는 크게 표준화된 기술규칙을 일관되게 적용하는 데이터의 생산과정에서의 질적 통제, 전거 파일을 통한 이름과 표제 접근점의 통제, 주제표목과 분류표의 개발, 기계가독 레코드의 규칙과 형식의 표준화 등에서 유래한다. 그 외에 도서관측에서의 목록 표준화는 부수적 가치를 갖는데, 이는 바로 도서관의 목록이 거대한 정보의 저장고에 대한 종합적 접근장치로서 효과적 기능을 수행할 수 있는 잠재성을 갖는다는 것이다.[20]

전자정보자원이 보편화되는 시대에도 목록이 도서관의 주된 서지 도구로서 자리 잡기 위해서는 먼저 컴퓨터와 네트워크 기술의 효율성을 도입하는 것이 필요하다는 것이다. 그리고 기존의 목록이 주로 인쇄매체 중심의 문헌을 대상으로 한 것이었다면 이제 목록규칙이나 MARC의 형식은 다양한 전자정보자원의 기술과 검색을 충분히 고려하여 조정되어야 한다.

전자정보자원의 목록도 전통적으로 목록이 갖는 다른 검색도구들과의 차별성과 구조적 완전성과 통일성, 기술의 정확성이 지속적으로 추구되어야 한다. 분류 기법이나 다양한 언어 통제에 의한 논리적이고 체계적인 검색의 방법론들, 서지 레코드 형식에 관한 정밀한 규칙과 표준화, 기술의 정확성과 통일성 등 전통적으로 목록이 갖고 있는 원칙과 특성은 목록으로 하여금 사회적 정보저장고에 대한 효과적 접근장치로서 뿐 아니라 정보자원의 효과적 통제도구로서의 역할을 할 수 있게 하는 것이다.

한편 목록 레코드를 작성하는데는 전문적인 지식과 시간이 많이 투입되어야 하므로 결국은 고비용이 문제가 될 수 있으며 이는 목록이 갖는 취약점일 수도 있다. 그러나 이 때문에 목록을 포기한다는 것은 설득력이 없다. 이 문제는 도서관계의 협력과 정보기술의 활용으로 해결할 수 있는 방안을 찾아야 할 것이다. 목록규칙과 MARC 형식의 표준을 강화하고 여기에 온라인 기술을 적극적으로 활용함으로써 훨씬 수월하게 목록레코드의 작성을 분담하고 이를 공유할 수 있게 되었다. 이와 같은 협력의 방식과 효과는 전자정보자원의 서지 레코드 작성에도 그대로 적용될 수 있으며 그 결과 개체 도서관들의 목록 부담과 비용을 획기적으로 절감할 수 있게 될 것이다. 도서관간의 협동에 의해 전자정보자원의 MARC를 작성하고 있는 예는 OCLC의 CORC(Cooperative Online Resource Catalog)[21]

---

20) Tom Delsey, The Library Catalogue in a Networked Environment, 〈http://lcweb.loc.gov/catdir/bibcontrol/delsey_paper.html〉 [cited 2004. 4. 22].

등에서 찾아볼 수 있다.

한편, 전자정보자원 서지 레코드의 분담과 공유에 관한 한 앞서 인용한 허쓰웨이트의 지적대로 '어떻게 목록 할 것인가' 이전에 오히려 '무엇을 목록 할 것인가' 하는 것이 선결 과제가 될 것이다 목록의 문제 이전에 질적으로 검증되지 못한 정보의 더미에서 보존의 가치를 지닌 정보자원을 선별하고 이를 도서관의 정보자원으로 축적 보존하는 방안이 마련되어야 한다는 것이다.

전자정보자원의 수집과 보존을 위하여 국가적 차원에서의 집중적 수집이나 계획적 분담이 절실히 요구된다. 전자정보자원의 현재적 상태나 팽창 추세를 감안할 때 개체 도서관들이 산발적으로 이들 자원을 수집 보존하는 방법으로는 도서관간의 불필요한 정보자원 중복으로 재원의 낭비를 가져올 수 있을 뿐 아니라 그 망라성을 보장할 수도 없기 때문이다.

미국 의회도서관에서는 국가적 차원의 전자정보자원 수집과 보존 프로그램인 The National Digital Information Infrastructure and Preservation Program[22] 을 수행하고 있으며 그 외에 세계 여러 나라에서도 이에 대한 관심을 갖거나 실행하고 있는 것으로 알려지고 있다.[23] 그러나 우리 나라에서는 아직 네트워크 상의 정보자원에 대한 체계적인 수집이나 편목에 관한 국가적 차원의 협력 체계가 실행되고 있지 못하다. 우리 나라도 전자정보자원의 체계적 수집과 보존을 위한 계획을 수립하는 것이 시급하다. 우리 나라의 네트워크 발달과 전자정보자원 증가 추세가 세계 여느 나라에 비교해도 뒤쳐지지는 않기 때문이다.

그 방법은 국가중앙도서관과 같은 특정기관이 집중적으로 수집 관리하거나, 다양한 분야의 중추적 기관들이 분담하여 수집하는 등의 방안이 있을 수 있다. 그러나 이와 같은 사업은 많은 시간과 인력, 비용이 소요되는 것이므로 국가적 차원의 지원과 계획 아래 가급적 여러 기관들이 분담하는 방안이 보다 효율적일 것으로 생각된다.

우리 나라에서도 이러한 협력의 가능성에 대하여 기대를 가질 수 있게 하는 것 중의 하나가 '국가전자도서관 구축사업'과 같은 프로젝트이다. 국가전자도서관의 웹사이트에서는 이 사업의 추진 배경과 효과에 대하여 다음과 같이 기술하고 있다.[24]

> 국가전자도서관 구축사업은 국내 주요 도서관을 연계하여 국가정보능력을 신장시키고 지역간 균형 있는 발전을 꾀하며, 향후 전국 도서관들을 연결하여 국가문헌정보 유통체제를 구축하는 정보화 기반을 확립하고자 …
>
> 가. 대 국민 정보서비스 측면
>
> - 국가전자도서관시스템 안정화 및 성능개선에 따른 대민 서비스 향상

---

21) CORC는 여러 도서관의 협력에 의해 메타데이터를 생산하고 공유하는 프로젝트로서 여기에는 한 시스템 안에 DC와 MARC가 통합되어 있다. BUBL LINK home page, 〈http://bubl.ac.uk/link〉 [cited 2005. 3. 15].

22) Digital preservation home page, 〈http://www.digitalpreservation.gov〉 [cited 2005. 3. 15].

23) 이재윤, 조현주, "인터넷자원의 보존 및 장기적 이용을 위한 국가프로젝트," 情報管理硏究, 제29권, 제3호(1998), p.38의 도표 참고.

24) 국가전자도서관 홈페이지, 〈http://www.dlibrary.go.kr/NEL/Introduct/business.jsp〉 [인용2004. 9. 1].

- 명실 상부한 지식기반사회 구현의 확고한 토대 마련
- 신속하고 신뢰성 있는 원문정보 서비스의 지원으로 국민의 정보획득 비용 및 노력 절감

나. 국가 경쟁력 제고 측면

- 기존 전자도서관 시스템의 적극 재활용과 국가자료 구축 및 DB 운영에 따른 비효율성 제거를 통한 예산 절감.
- 국가자료(원문, 목차, 초록 등)의 중복 구축을 최소화하여 효율성 및 경제성 제고.
- 도서관간 정보공유 체제 강화를 통한 중복투자 방지
- 명실상부한 전자도서관 표준 모델 확립을 통한 국내 타 전자도서관 구축의 가이드라인 제시
- 저작권 자료 이용 및 보호의 안정적 토대 마련을 통한 저작물 유통의 활성화에 기여
- 국가 자료의 일관되고 안정화된 서비스 및 관리체계확립을 통해 지식기반 사회의 조기확립에 기여

그리고 국가전자도서관 체제[25]에 따르면 국립중앙도서관과 국회도서관, 법원도서관, 한국과학기술정보연구원, 한국교육학술정보원, KAIST 과학도서관, 농업진흥청 농업과학도서관 등 6개의 공동참여기관이 참여하고 있는 것으로 나타나고 있으며 현재 상태로는 여기에 참여한 기관들이 독자적으로 구축한 데이터 베이스를 각각 검색 이용할 수 있도록 링크 한 수준임을 볼 수 있다.

이러한 국가적 계획이 제대로 성과를 거두기 위해서는 먼저 참여 기관들의 담당 분야를 철저하게 구분함으로써 기관간 정보자원의 중복을 최소화하고 역으로 이들의 분담 분야를 통합하면 누락되는 부분이 없이 전 학문 분야를 아우를 수 있도록 조정하여야 한다. 그리고 여기에 전자정보자원까지 포함하게 한다면 국가적 정보자원의 망라적 수집과 보존, 공유를 실현할 수 있게 될 것이다. 물론 여기에는 전자정보자원의 목록 작업과 공동 이용의 문제도 부수되어야 함은 당연할 것이다.

이러한 계획의 실행에는 부수되는 문제들도 있을 수 있다. 전자정보자원의 복잡한 저작권 문제를 어떻게 해결할 것인가 하는 것이나, 클리포드 스톨[26]이 신랄하게 지적했듯이 이러한 매체들을 이용하거나 읽을 수 있는 소프트웨어의 확보와 장기적 보존 등도 문제가 될 수 있을 것이다.

Ⅵ. 마치는 글

목록업무에 컴퓨터의 도입은 MARC를 탄생시킴으로써 그 형태와 내용의 획기적 변화를 가져왔

25) 상게 웹사이트.

26) 클리포드 스톨 지음, 허풍떠는 인터넷, 한경훈 옮김(서울 : 세종서적, 1996), pp.258-259.
전자문서를 제작하거나 이용하는 소프트웨어는 빠른 속도로 개정된다. 이러한 과정에서 보존되는 전자문서들 중에는 세월이 지나면서 이를 읽을 수 있는 소프트웨어가 없어서 문서로서의 구실을 할 수 없는 경우도 생긴다. 이는 도스(DOS)가 서서히 사라져 가는 현상이나, 낮은 버전의 소프트웨어에서 운영되던 프로그램들이 새로운 버전에서는 실행되지 않는 현상들을 자주 접하면서 고민하지 않으면 안될 문제가 되고 있다.

다. 나아가 전자정보자원이 점차 늘어나고 특히 네트워크를 통해 빠르게 유통되는 정보들이 늘어나면서 이러한 정보자원들을 쉽고 빠르게 조직하고 검색할 수 있도록 하는 서지 도구로써 검색엔진, 메타데이터와 같은 것들에 대한 논의가 빈번해지고 있다. 이와 같이 메타데이터가 MARC와 같은 목록의 대안으로 논의되고 있는 것은 구조적 복잡성과 고비용의 문제를 갖고 있는 목록에 비하여 신속하게 기계적으로 서지 레코드를 작성할 수 있다는 것이 그 이유 중 하나이다.

그러나 도서관의 입장에서 이러한 검색엔진이나 메타데이터와 같은 것들이 전통적 목록의 역할을 온전히 대체할 수 있을 것인지에 대하여는 심각하게 재고해 보지 않을 수 없다. 본고에서는 도서관의 서지 도구로써 인터넷 검색엔진이나 메타데이터가 갖는 문제점들과 이러한 문제점을 극복하기 위한 대안으로 전통적 목록이 지향해야 할 방향에 대하여 고찰하였다.

도서관의 목록은 정보자원의 신속한 검색이나 조직 차원을 넘어 완전하고 정확한 기술, 엄격한 표준화를 통한 정보자원의 공유와 체계적 서지통정이라는 목적을 달성해야 하는 종합적 접근장치이자 도서관 관리의 핵심 서지 도구이다. 따라서 이와 같은 서지 도구가 일차적으로 정보 생산자의 자의적 판단에 의해 작성된 데이터에 의존하여 만들어진다는 것은 용납되기 어렵다. 이는 책의 저자로 하여금 목록을 작성하게 하는 것과 크게 다르지 않기 때문이다. 도서관의 서지 레코드는 정보자원을 수집 관리하는 기관에 의해 완전하고 정확하게 기술되어야 하며 통일된 규정에 따라 작성됨으로써 표준화를 기할 수 있고 통합과 공유의 효과도 극대화할 수 있다. 그러므로 도서관의 입장에서는 전자정보자원에 대하여도 여전히 MARC와 같은 목록의 작성과 유지가 강조되어야 할 것이다.

기존의 목록이 갖는 구조적 복잡성과 고비용 문제는 도서관 관리의 장애 요소가 될 수 있다. 그러나 목록의 구조적 문제는 목록이 인류의 거대한 정보 저장고에 대한 종합적 접근장치로서 완전성을 유지하기 위한 불가피한 선택이다. 따라서 이러한 문제는 피해가기 보다는 극복하기 위한 도서관계의 공동 노력이 절실히 요구되는 것이다.

전자정보자원의 옥석을 구분하고 보존을 위한 정보자원으로 선별하는 일, 그리고 이들에 대한 서지 도구를 작성하는 일에 도서관계가 체계적으로 분담 협력함으로써 도서관간의 정보자원 관리의 중복을 줄이고 개체도서관들의 목록 업무 부담을 경감할 수 있을 것이다.

우리 나라에서 전자정보자원의 체계적 수집과 조직 보존을 위하여 기왕에 발족한 국가전자도서관 구축사업과 같은 계획이 그 대안이 될 수도 있을 것이다. 이를 위해서는 사업의 범위를 전자정보자원의 수집과 보존, 그리고 목록 업무에까지 확대하고 이를 체계적으로 분담 공유할 수 있는 국가적 차원의 관리가 필요할 것이다.

〈참고문헌은 각주로 대신함〉

KCI