

디지털 도서관 아카이빙 현황 및 발전과제

국립중앙도서관 이재선 과장

1. 머리말

디지털도서관에서 아카이빙은 현재만이 아니라 후대의 세대를 위하여 디지털자원의 장기적인 접근이 가능하도록 다양한 디지털파일의 수집, 정리, 이용, 보존까지의 일련의 과정이라고 할 수 있다. 이러한 과정에서 디지털자원의 수집 및 보존을 위한 법적 기틀과 수행계획, 기술현황의 파악, 서비스 기준, 각종 장비에 대한 이해, 서비스를 지원하는 유저인터페이스와 스마트기기 등 상당한 기술에 대한 이해를 요한다. 특히 디지털자원의 특성은 디지털자원이 수록된 매체가 수명이 짧아 콘텐츠의 이용도 수명이 짧아진다는 점이다. 따라서 안전하게 보존하여 장기적인 이용을 보장한다는 것은 매체에 적용되는 기술의 변화를 늘 파악하고 대처할 수 있어야 한다는 점에서 기술에 대한 R&D가 중시되는 면이 있다.

웹기술이 발전함에 따라 디지털도서관에서는 소장 자료에 대한 디지털화 사업과 병행하여 웹을 통해 유통되는 문헌과 웹사이트 자체를 도서관 자료로 포함하여 수집하기 시작하였다. 이들 디지털자원에 대하여 특히 웹 아카이빙 기술을 사용하고 있으며 자체의 웹아카이브를 운영하고 있다. 이러한 활동은 국제도서관연맹(IFLA)과 세계국립도서관장회의(CDNL)와 같은 국제 도서관계 내에서 디지털 아카이빙에서의 표준 이용과 발전에 관한 활동들이 전개되어 왔다. 본고에서는 주요 국립도서관의 디지털도서관에서 운영되고 있는 아카이빙에 관한 현황과 과제를 소개하여 국내 관계자들과 함께 생각하는 토론의 시간을 갖고자 한다.

2. 디지털 아카이빙의 정의

디지털 아카이빙은 분야에 따라 상당히 다르게 사용되고 있다. 모든 디지털자료는 이용을 위해 특정 하드웨어와 소프트웨어를 필요로 하기 때문에 상당히 기계 의존적이다. 따라서 자료를 유지관리하기 위해 요구되는 측면도 행정적인 측면, 기술적인 측면, 장비적인 측면 등 매우 다양하다. 또한 기술의 발전으로 새로운 측면을 구사하기 위해 다양한 기법들이 나타나고 그에 따른 용어도 등장하고 있다. 최근에는 '디지털 큐레이션' 개념이 사용되고 있다.

아카이빙과 보존에 대한 용어 관계는 2006년 IFLA에서 발표한 주요 국립도서관의 디지털 보존 관련 조사결과에서 찾아 볼 수 있으며 그 내용은 다음과 같다.

디지털 보존 혹은 장기 보존은 스토리지와 접근 모두와 관련하여 디지털 혹은 전자 자료의 유지, 보호/큐레이션과 관련된 모든 활동에 대한 일반적인 용어이다. 장기간은 5년 이상을, 단기간은 5년 이내를 의미한다. 디지털보존에서 주된 활동은 디지털 아카이빙과 영구적인 접근으로 구분된다.

디지털 아카이빙은 디지털 자료의 백업과 지속적인 관리를 위한 과정을 의미한다. 이는 디지털 보존을 위한 전략을 의미하는 것은 아니다.¹⁾

여기에서 디지털 전략(strategy)이란 마이그레이션과 에뮬레이션과 같이 영구적으로 접근할 수 있도록 자료를 저장하고 유지하기 위한 방법에 관한 것으로 단순히 저장된 디지털 자료를 제공하는 것 이상의 모든 기술을 말한다.

1) Ingeborg Verheul [IFLA]. Networking for digital preservation: current practice in 15 national libraries. Munchen : Saur, 2006, p. 21. <http://www.ifla.org/files/hq/publications/ifla-publications-series-119.pdf>

한편 디지털 큐레이션 분야의 연구보고서 “e-Science Curation report” 에서의 디지털 아카이빙과 보존의 정의를 살펴보면 다음과 같다.

큐레이션은 데이터 작성의 관점에서 데이터의 이용을 관리하고 촉진하는 데 있어서 동 시대의 목적에 부합하고 데이터의 발견과 재이용이 가능하도록 하는 활동을 말한다. 동적인 데이터세트에 대해서는 그 용도에 맞게 지속적이고 참신성을 유지할 수 있도록 하며, 해설과 다른 출판된 자료와 연결 지어 차원 높은 큐레이션이 포함되는 것을 의미한다.

아카이빙은 데이터를 접근할 수 있도록 적절하게 선정·저장하고, 장기간 데이터의 안전성과 진본성을 포함하여 논리적·물리적 무결성이 보장되도록 하는 큐레이션 활동이다.

보존은 데이터의 특정 자료가 장기간 유지 관리되어 기술의 변천을 통해 지속적으로 접근 및 전달될 수 있도록 하는 아카이빙 내에서의 활동이다.²⁾

아카이빙에 대하여 위의 두 가지 정의는 서로 상반된 견해를 보인다. IFLA의 관점은 아카이빙을 보존활동 내에 속한다고 보고, 디지털 큐레이션의 관점은 아카이빙을 아키비스트의 활동을 강조하면서 아카이빙을 보존 보다 큰 개념으로 언급하였다. 그러나 미래 세대를 위해 장기간 보존해야 한다는 점에서는 동일하다.

디지털 큐레이션은 비교적 최근 나타난 용어로 데이터 아카이빙, 디지털 보존, 자원의 라이프 사이클 전반에 걸쳐 데이터의 실질적 관리와 평가를 아우르는 활동으로 아카이빙과 보존보다 상위의 개념이라고 할 수 있다. 디지털큐레이션센터(DCC)에서는 개념화, 생성, 수집에서 보존, 액세스, 매체 변천에 이르기까지의 라이프 사이클을 파악하여 데이터를 보존,

관리하는 ‘큐레이션’ 의 과정을 모형으로 작성하여 제공하고 있다.

이와 같은 용어들이 비슷한 개념으로 혼재되어 있는 것은 사용 분야에 따라 각각 강조하는 분야가 다른 데서 기인한다. IFLA에서 정의한 내용을 보면 도서관에서 아카이빙은 아날로그 자료를 디지털화하는 작업으로 볼 수 있고 장기 접근을 포함하면 보존으로 보고 있다. 도서관계는 디지털 보존이란 용어를 보편적으로 사용하는 것으로 생각할 수 있다. 반면에 디지털큐레이션센터(Digital Curation Center)의 정의는 한 차원 더 높여 디지털 자원에 해설이나 주석을 연결하거나 관련 다른 출판물과 연결을 포함한다. 마치 미술관의 큐레이터의 해설이 곁들여지는 것처럼 디지털자원을 해석할 수 있는 풍부한 자료를 연결하여 제공하는 것까지를 포함하는 것을 의미한다.

도서관에서 아카이빙 대상은 디지털화한 소장자료의 디지털파일, 전자책이나 전자저널, 동영상 등 온라인의 디지털 파일, 웹사이트 등을 들 수 있다. 이러한 자료들은 아카이빙을 통해 구현되고 디지털도서관을 통해 서비스된다.

디지털도서관에서 실제 아카이빙을 위해서는 법적 토대 마련, 디지털정보의 수집과 장서개발, 저작권문제, 식별과 목록작성, 보존스토리지, 보존, 접근과 같은 일련의 과정이 필요하나 본고에서는 지면관계상 보존문제는 논외로 하고자 한다.

2) Philip Lord and Alison Macdonald e-Science Curation Report, 2003.
http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf,
Elizabeth Yakel, Digital curation, OCLC Systems & Services, Volume 23, Number 4, 2007, pp. 335-340(6)

3. 디지털도서관의 아카이빙 현황

3.1. 온라인 자료의 수집과 관련 법령

주요 국립도서관의 온라인자료 수집현황을 한 눈에 알 수 있는 조사결과가 2011년 세계국립도서관장 회의에서 발표되었다. 영국 국립도서관 주관으로 실시된 조사에는 55개국의 국립도서관이 회신을 보내왔다. 그 결과에 따르면 최소한 오프라인의 전자자료에 대한 납본규정을 가지고 있는 도서관은 35개관이었고, 전자책, 전자저널 및 기타 상업적인 온라인 출판물을 포함하여 온라인자료에 대한 수집이 가능한 도서관은 이 중 35%를 차지하였다.

웹 아카이빙을 위한 허베스팅이 가능한 도서관은 30여 곳이었다. 미국의회도서관은 인터넷 아카이브(Internet Archive)에 의해 수행되고 있어서 노르웨이에서는 개인정보 문제 때문에 특별한 신청에 의해서만 접근이 허락된다. 프랑스와 독일 국립도서관의 경우는 국립도서관 이용자로 등록된 경우에 한하여 관내에서만 이용할 수 있다. 영국의 경우는 동일 웹 사이트에 대해 한 번에 한 명의 이용자로 규정상 접근이 제한되어 있다. 오스트리아와 캐나다, 핀란드에서는 국립도서관 내에서 누구나 이용할 수 있다. 호주와 일본, 싱가포르 등은 이용이 허락된 웹사이트에 대해서는 온라인으로 아카이브를 자유롭게 이용할 수 있다.

전자책과 전자저널 수집은 약 30개 국립도서관이 관련 법규를 가지고 있다. 그러나 이 가운데 납본을 통해 수집하는 전자책과 전자저널의 경우는 19개 국립도서관에 불과하다. 네덜란드 국립도서관의 경우 전자저널 출판사를 위해 아카이빙 서비스를 제공해왔

다. 영국의 경우 전자저널에 대해 자발적인 납본 방식을 취하고 있고, 스위스는 전자저널과 전자책 모두를 자발적 납본에 의존하고 있다.

네덜란드국립도서관 등 7개 국립도서관은 상업적으로 출판된 저작에 대해 보상금 없이 오픈 액세스 차원에서 수집하고 있다. 5개 국립도서관은 오픈 액세스는 아니지만 상업적으로 출판된 자료를 수집하고 있다. 11개 국립도서관은 두 가지 방식을 모두 취하고 있다. 26개국에서 관련 법에 의해 혹은 자발적인 방식을 통해 수집하고 있다.³⁾

온라인자료에 대한 납본법을 가장 먼저 규정한 나라는 중국으로 1996년 전자출판물에 관한 수집을 국가 저작권기관에 고지함으로써 시작되었다. 캐나다의 경우는 2004년 캐나다 도서관문서관법에서 규정하고 있다. 스웨덴은 2002년 웹 페이지와 온라인 데이터베이스의 수집에 관해 정부와 특별 협약을 맺으면서 웹아카이빙이 시작되었다.

장기 보존과 접근은 원본의 사본작성 없이는 불가능하다. 온라인자료는 저작권법에 의해 보호받으며 복제를 막기 위해 기술적인 제한이 자주 사용된다. 적어도 장기 보존과 접근을 위해서는 특수한 환경임을 부분적으로 허용해야 한다. 경우에 따라서는 해당 분야의 단체나 온라인자료의 소유권자와 별도의 계약이 필요하다.

금년 헬싱키에서 개최된 세계국립도서관장회의에서는 온라인자료를 포함한 전자출판물 납본을 강화하기 위해 “e-Legal Deposit Advocacy Pack”에 대한 논의가 있었다. 그 배경에는 웹 페이지가 평균 75일 만에 사라지는 환경을 따라가지 못하는 법 제정의 지연에서 비롯되었다. 이에 회의 참가자들은 2003년 발표된 유네스코 ‘디지털 유산 보존 헌장’을 재차 강조하며 온라인자료의 “블랙홀”이 생

3) British Library international survey on E-Legal Deposit 2011: Summary of findings
[http://www.cdnl.info/2011/pdf/e_2Dlegaldeposit_20survey_20CDNL_20Slides_20Aug%20\[Compatibility%20Mode\].pdf](http://www.cdnl.info/2011/pdf/e_2Dlegaldeposit_20survey_20CDNL_20Slides_20Aug%20[Compatibility%20Mode].pdf)

기지 않도록 온라인자료 수집을 포함하는 납본법 개정의 확산에 동참할 것을 촉구하였다.

국립중앙도서관은 2009년 9월 26일에 개정된 도서관법에 따라 국내에서 생산된 온라인자료를 선별적으로 수집하고 정당한 보상을 하도록 되어있다. 이를 근거로 2010년부터는 본격적으로 국내의 온라인 자료를 수집하고 있다. 이에 의거하여 2011년말 현재는 전자책 37,000책, 전자저널 약 2,800종, 이미지 등 기타자료 80,000여 건을 수집하였으며, 웹 페이지 60,000여건, 웹문헌 650,000건을 구축하였다.

3.2. 아카이빙 시스템

웹 아카이빙의 업무 흐름은 일반적인 도서관 업무와 유사하다. 웹자원을 선정하고, 수집, 조직, 저장, 메타데이터 작성 및 접근의 과정으로 진행된다.

3.2.1. 선정 및 수집

웹자원의 방대함과 예산의 한계로 해서 디지털자원은 주로 선정되고 있다. 그러나 전술한 바와 같이 웹상의 자국내 도메인을 모두 수집하는 경우도 있다. 전량을 수집하는 것은 웹 크롤러 로봇에 전적으로 의존하고 메타데이터를 자동으로 추출하는 기법을 사용한다. 실무자의 입장에서 이 방식은 오히려 간단한 공정을 요하므로 편리한 면이 있다. 선택적인 수집을 취하는 경우, 특정 이슈를 대상으로 수집한다. 미국 의회도서관은 선택적 수집으로 9.11사태, 선거, 이라크전쟁 등 관련 사이트뿐만 유튜브 상의 비디오도 수집하였다. 캘리포니아 디지털도서관의 경우, 로스앤젤레스에서 발행한 산불 관련 등을 수집하였다. 일부에서는 블로그, 신문, 버추얼 세계 등을 수집하기도 한다. 프랑스 국립도서관의 경우, 자국 내 도메인 전량과 선정에 의한 수집의 두 가지 방식을 모두 사용한다. 자료의 유형으로는 e-diary와 같은 자료도 수집한다. 체코국립도서관의 경우, 체코 도메인 “.cz”

이외의 in .net, .com, .org, or .edu 도메인을 포함하여 자동으로 수집한다.

웹사이트를 인력이 선정하는 경우에는 상당한 시간과 노력이 필요하다. 웹사이트를 선정하려면 사이트를 평가해야 하고 일정 기준을 고려해야 하며, 특정 주제에 따라 일관성 있게 수집해야 하기 때문이다. 대부분의 경우 주제별로 서비스 담당 사서들이 참여하여 웹사이트를 선정한다. 미국 의회도서관의 경우 웹사이트 선정에 80여명의 사서가 참여하며 이들이 선정한 사이트를 로봇에 심어 하베스팅한다. 일부 도서관에서는 사이트 선정에 드는 비용을 줄이기 위해 이용자가 추천한 URL을 받아들이기도 한다.

국립중앙도서관에서는 선택적 수집을 하고 있으며 주로 정부기관과 공공기관, 연구 학술단체, 법인 또는 재단 등 각종 사회단체, 선거 관련 사이트 등 국가적 이슈를 중심으로 수집하고 있다. 선정에는 사서와 외부전문가들이 참여하고 있다. 선정대상이 되는 웹사이트에 대해서는 해당 기관에 사용허락을 요청하여 동의한 경우에 한하여 공개한다.

웹아카이빙에서 수집은 로봇이 해당 사이트를 방문하여 콘텐츠를 가져오으로써 이루어진다. 크롤러가 해당사이트를 방문하려면 해당 서버에 URL을 입력하여야 한다. 입력된 URL을 따라 웹사이트를 찾아갈 수 있다. 웹 크롤러 로봇을 사용하기 위해 다수의 국가에서 국제인터넷보존컨소시엄(IIPC)이 제공하는 오픈 소스 Heritrix 웹 크롤러 소프트웨어를 사용한다. 유저 인터페이스에는 인터넷 아카이브에서 제공하는 Wayback Machine을 사용하며 검색 엔진을 위해 NutchWax, Solr 등 다양한 소프트웨어를 사용한다. 공개 소프트웨어는 그대로 사용하기 보다는 자국의 사정과 언어에 적합하도록 개발되어야 한다. 영국 국립도서관의 경우에는 상용제품인 DigiTool을 사용하기도 하고 네덜란드국립도서관과 같이 IBM과 공동으로 자관에 필요한 기능을 확장하여 자체 개발

하기도 한다. 네덜란드의 Digital Information Archival System(DIAS)은 IBM이 상용화하려 하였지만 널리 보급되지는 않았다. 현재 독일국립도서관은 네덜란드의 DAIS의 일부 기능만을 도입하여 사용하고 있다. 디지털 웹 아카이빙을 완전하게 구사하여 운영하는 도서관은 호주, 네덜란드, 오스트리아, 뉴질랜드 등 일부의 국가에서 이루어지고 있으며 미국 의회도서관과 NDIIPP에 참여하는 도서관이 aDORe와 같은 시스템을 테스트한 바 있으며, 현재는 LCWA(Library of Congress Web Archive)를 구축하여 운영하고 있다.

웹아카이빙 시스템을 개발할 때 표준으로 Open Archival Information System(OAIS) 모형을 참고한다. 이는 CCSDS(Council of the Consultative Committee for Space Data System)에 기원을 두고 있다. OAIS 표준은 2003년 ISO 14721로 제정되었다.

국립중앙도서관은 2004년부터 웹아카이빙을 위해 OASIS(Online Archiving and Searching Internet Sources) 사이트를 구축하여 운영하고 있다. 수집된 웹사이트와 웹문헌은 디지털장서관리시스템에 저장된다. 저장된 파일을 기초로 메타데이터를 작성하고 동 시스템을 통해 디브리리 포털로 검색에 제공된다. 수집 로봇으로 Heritrix 웹 크롤러 소프트웨어를 사용하며 검색 인터페이스를 위해 Wayback Machine, 메타데이터 등의 색인 작성을 위해 Solr 엔진 등을 사용하고 있다.

3.2.2. 수집 자료의 파일 형식

도서관이 수집하는 자료는 모든 형식의 파일을 수집한다. 이미지용 TIFF, 텍스트에 대하여 PDF와 XML, 오디오에 대하여 WAV, 웹사이트에 대하여 HTML 등 실로 다양하다. 자주 사용되는 형식은 ZIP, PNG, ISO, JPEG, GIF, TeXm PS, EPS, AI, RTF, TXT 및 ASCII 등이다. 최근에는 보존용

으로 PDF/A와 ePub 등이 ISO 규격으로 등재되면서 그 사용이 확산되고 있다.

국립중앙도서관은 전자책과 전자저널에 대해 PDF/A 등의 파일로 수집하고 있으며, 기타 데이터베이스 구축기관에서 사용한 파일들도 수집하고 있다. 이 밖에도 오디오와 비디오 파일, 장애인을 위한 데이지 파일 등이 있다.

3.2.3. 메타데이터 작성

메타데이터 형식은 디지털 자원의 이용을 위해 다양하게 개발되었다. 데이터의 형식은 더블링크어 메타데이터 형식, LC의 주제명표 LCSH, 각각의 기관이 작성한 전거데이터를 적용하기도 한다. 자동으로 수집되지 않아 화면을 캡처한 경우에는 사람이 직접 화면을 보면서 메타데이터를 작성한다. 주로 사용되는 메타데이터 형식에는 데이터 관리용으로 Metadata Encoding and Transmission Standard(METS), 디지털자원에 대한 기술 형식으로 Metadata Object Description Schema(MODS), 전거용 메타데이터 Metadata Authority Description Schema(MADS), 이미지용 메타데이터 Metadata for Images (MIX), 보존 정보를 위해 PREMIS가 개발되어 있다. 이 외에도 기존 MARC 데이터의 활용을 위해 MARC/XML, 더블링크어 메타데이터(Dublin Core Metadata)이 개발되어 있다. 이러한 표준들은 미국의회도서관 네트워크개발 및 MARC 표준실(Network Development and MARC Standards Office)에서 관리 운영하고 있다.

국립중앙도서관에서는 수집되는 온라인자료에 대해 디지털장서관리시스템 개발 당시 MODS, METS, PREMIS를 적용하였으며, 본격적으로 데이터를 작성한 것은 2010년부터이다. 외부기관이 작성한 디지털 파일의 연계에는 더블링크어 심플 버전을 사용하고 있다.

3.2.4. 접근과 이용

디지털 자료에 대한 접근은 저작권의 제한 등으로 관내 서비스와 관외 서비스로 구분된다. 고서와 고문서, 공공기관의 간행물, 저작권자의 동의를 얻은 자료에 대해서는 관외에서도 이용가능하다. 저작권이 존재하는 경우에는 관내에서만 이용가능하다.

네덜란드국립도서관은 2005년 여름부터 오픈 액세스에 대하여 온라인 으로 제공하고 있으며, 스웨덴의 경우 상업적 출판물과 비상업적 출판물을 구분하여 제공하고 있다. 프랑스국립도서관에서는 금년부터 저작권이 존재하는 전자책에 대해서도 저작권집중관리 기구를 통한 서비스를 계획하고 있다.

대부분의 웹아카이브는 관내에서 주로 이용되는 다크 아카이브로 운영되고 있다. 일부 도서관은 공개적으로 웹 아카이브를 운영하는 경우에도 웹사이트 소유자와의 충돌을 피하기 위해 접근 기능을 떨어뜨리거나 지연시킨다. 예를 들면 웹사이트를 수집한 후 최소 3개월을 지연시키기도 하며, IA에서 새로 시작한 Archive-it의 경우에는 6-12개월을 지연시키기도 한다. 검색 능력은 미국 의회도서관이나 뉴질랜드 국립도서관의 경우, 선거통제를 거쳐 검색결과를 보여줌으로써 상당히 품질 좋은 정보를 제공한다. 이 두 기관에서는 사람이 직접 데이터를 보면서 주제명을 연결하는 작업을 한다. Wayback Machine에 기초한 웹아카이브는 URL에 의해서만 검색이 가능한 반면, NutchWax 검색 엔진은 전문 검색을 지원할 수 있다. 발전된 웹아카이브로 콘텐츠 마이닝, 태그 클라우드, 3D wall 기법을 도입한 영국 국립도서관의 UK Web Archive를 들 수 있다.

국립중앙도서관의 경우, 소장자료를 디지털화한 원문정보는 약 420,000건으로 이 가운데 150,000건은 저작권이 소멸되었거나 이용허락을 받아 디브리리 포털을 통해 인터넷 서비스를 하고 있다. 저작권이 존재하는 27만건에 대해서는 디지털도서관 정보광장에

서 서비스되고 있다. OASIS의 경우, 총 구축건수 710,000건 중 공개적으로 이용할 수 있는 데이터는 74,000 건이다. 대부분은 관내에서 오아시스 홈페이지를 통해 서비스되고 있다. 웹 문헌에 대해서는 디브리리 포털에서 통합검색으로 제공된다.

4. 국제 협력 활동

4.1. Internet Archive

인터넷 아카이브(IA)는 1996년 설립된 미국의 비영리법인으로 디지털 포맷으로 존재하는 역사적인 컬렉션의 영구적인 접근을 제공을 목적으로 한다. IA는 전 세계를 대상으로 웹정보를 수집하고 있다. 관리자가 액세스를 제한하고 있는 경우를 제외하고는 인터넷에서 공개되어 있는 정보 모두를 수집대상으로 한다. 현재는 웹페이지뿐만 아니라 텍스트, 오디오, 동영상자료, 소프트웨어까지 포함하여 제공하고 있다.

IA의 데이터베이스 규모는 인터넷의 확대와 함께 성장하여 매월 압축파일로 100 테라바이트에 가까운 용량이 증가하고 있다. 이러한 정보는 “Wayback Machine” 이라는 시스템을 통해 인터넷에서 이용할 수 있다. 또 IA에서는 “Archive-It” 라는 유료 서비스를 제공하고 있다. 이 서비스는 전문적인 지식이 없는 기관이라도 자관의 웹사이트 아카이브를 만들 수 있게 해준다. 이 서비스 계약자가 지정한 데이터를 IA가 수집하여 메타데이터를 작성하고 전문검색이 가능하도록 한다. 미국 정부기관이나 대학 등 다양한 기관이 이 서비스를 이용하여 자관의 웹사이트를 아카이브하고 있다.

그동안 IA는 도서에 대한 스캔작업을 추진하여 왔는데 2010년에는 그 양이 100만책을 넘어서었다. 2011년 2월에는 공공도서관 150개관과의 협력을 통해 IA가 운영하는 디지털화 도서 공개사이트

“OpenLibrary.Org”가 8만종 이상의 전자책을 참가관의 관내에서 열람할 수 있는 서비스를 개시하였다. OpenLibrary.Org의 도서를 이용하려면 Adobe Digital Editions를 통해서 IA가 제공하는 브라우저용 포맷 혹은 PDF, ePub 포맷으로 대출하여 자신의 PC나 iPad와 같은 디바이스나 도서관 내 컴퓨터에서 열람할 수 있다. 1회에 5책을 2주간 빌릴 수 있다. 또한 2011년 6월에는 물리적공간의 아카이브를 설치하여 온도관리가 되는 밀집형 수장시설을 갖추고 도서관 레코드, 영화자료를 보존하고 있다. 도서 30만 책을 우선 보존하고 있다.

2012년 8월부터 토렌트(Torrent) 형식으로 공공도메인의 영화나 라이브 콘서트, 라디오 프로그램, 오디오북, 도서 등을 140만 건 이상 제공하고 있다. 앞으로 BitTorrent를 이용해 제공할 뿐만 아니라 분산보존 시스템으로 활용할 계획을 가지고 있다.

4.2. IIPC

국제인터넷보존컨소시엄 (International Internet Preservation Consortium)은 웹 아카이브의 국제 협력을 목적으로 2003년 결성되었다. 이 컨소시엄에는 주요국의 국립도서관 등 40개 이상의 기관이 참여하고 있다. 국립중앙도서관은 2008년부터 회원으로 활동하고 있다.

IIPC는 크롤러 개발과 데이터 저장 형식과 같은 규격의 표준화, 검색 시스템의 검토, 웹 아카이브의 다양한 과제에 대해 협력하며, 정기적으로 회의를 개최하고 각 기관의 웹 아카이브의 사례와 개발 성과를 공유하고 있다. 매년 정기적으로 회의를 개최하며 하베스팅, 액세스, 보존 분야의 각각의 워킹그룹을 통해 실무에서 일어나는 개선사항을 공유하고 공동연구 개발을 수행하기도 한다. 연구 결과는 공개되어 오픈소스로 제공되며 웹아카이브에 필요한 기술들을 지원한다.

5. 맺음말과 디지털도서관 아카이빙의 과제

디지털 아카이빙 기술은 날로 발전하고 있고 디지털 정보는 무한히 증가하고 있다. 새로운 기술의 발전이 지속된다고 하더라도 모든 자료를 수집하고 조직하는 것은 불가능한 일이다. 도서관 간의 나아가 도서관과 유사한 기능을 하는 문서관과 박물관 간의 협력이 보다 활발하게 진행될 필요가 있다. 법적 제도적으로 확고히 하고 웹크롤러의 성능을 개선하고 메타데이터의 표준을 안정적으로 적용하여 상호 이용성을 높이고, 보존체계를 확립하는 모든 활동에서 항상 해결해야할 과제와 대면하게 된다. 이를 가장 효과적으로 해결하는 방법은 유관기관의 공조체제를 다지고 협력을 통한 공동대처가 최선의 방법이라고 생각한다.

소장자료의 디지털화 사업은 비교적 안정적인 반면에 웹아카이빙에 대한 작업은 웹환경의 변화로 늘 새로운 기술이 필요하다. 한국의 웹사이트는 웹사이트 작성의 원칙을 지키지 않는 경우가 많으며 텍스트 보다는 동적 이미지를 선호하며 암호사용을 통해 잠금장치가 다수 사용되어 아카이빙의 난제가 되고 있다. 인터넷에는 국경이 없다. 웹아카이브를 발전시키기 위해서는 법제도 개선과 아카이빙 기술 개발이 중요하며 국내외의 유관기관과의 협력이 필요하다. 국립중앙도서관은 영국국립도서관이 도입한 새로운 기법과 IA-it와 같은 해외의 우수사례를 참고하여 향후 오아시스 사업 환경을 개선하고 보다 발전된 서비스를 제공하기 위해 노력하는 한편, 국제적인 기술개발에 적극 참여하기 위해 힘쓸 것이다.

예로부터 아카이브란은 문서의 보존기능을 중시하였으며, 문서의 보관 및 복원에 관한 기술을 발전시켜 왔다. 앞으로 이러한 기술을 도서관과 협력하여

공유함으로써 양 기관의 발전을 도모하고 이용자에 대한 서비스 품질을 제고하는 일에 더욱 노력해 나가야 할 것이다.

[참고문헌]

1. Ingeborg Verheul [IFLA]. Networking for digital preservation: current practice in 15 national libraries, Munchen : Saur, 2006, p. 21.

<http://www.ifla.org/files/hq/publications/ifla-publications-series-119.pdf>

2. Philip Lord and Alison Macdonald e-Science Curation Report, 2003.

http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf.

3. Elizabeth Yakel, Digital curation, OCLC Systems & Services, Volume 23, Number 4, 2007 , pp. 335-340(6)

4. 기타 국립도서관 및 웹아카이브 홈페이지

