인터넷자원의 보존 및 장기적 이용을 위한 국가 프로젝트

National Projects for Preservation and Long-term Access to Internet Resources

이 재 윤*, 조 현 주**
(Jae-Yun Lee, Hyun-Joo Cho)

초 록

인터넷상의 디지털 문헌이 증가하고 디지털 도서관이 개발되면서 인터넷 자원을 국가 유산으로 간주하여 국가 차원에서 개발, 보존, 이용시키는 정책이 필요하게 되었다. 이에 본 연구에서는 외국의 사례를 살펴보고 우리 나라에 맞는 모형을 모색해보았다.

키 워 드

인터넷 자원, 보존, 디지털 도서관, 아카이브

ABSTRACT

Responding to the growing trend of Internet information resources, national policies to develop, preserve, and supply those resources are needed. The purpose of this paper is mainly to review and analyze some cases of foreign countries, and to build a model for national policy on internet archive in Korea.

KEYWORDS

Internet Resources, Preservation, Digital Library, Archive

* 연세대학교 문헌정보학과 강사
(Lecturer, Dept. of Lib. and Info. Science, Yonsei Univ.)
** 연세대학교 대학원 문헌정보학과 석사과정
(Master Course Student, Dept. of Lib. and Info. Science, Graduate School of Yonsei Univ.)

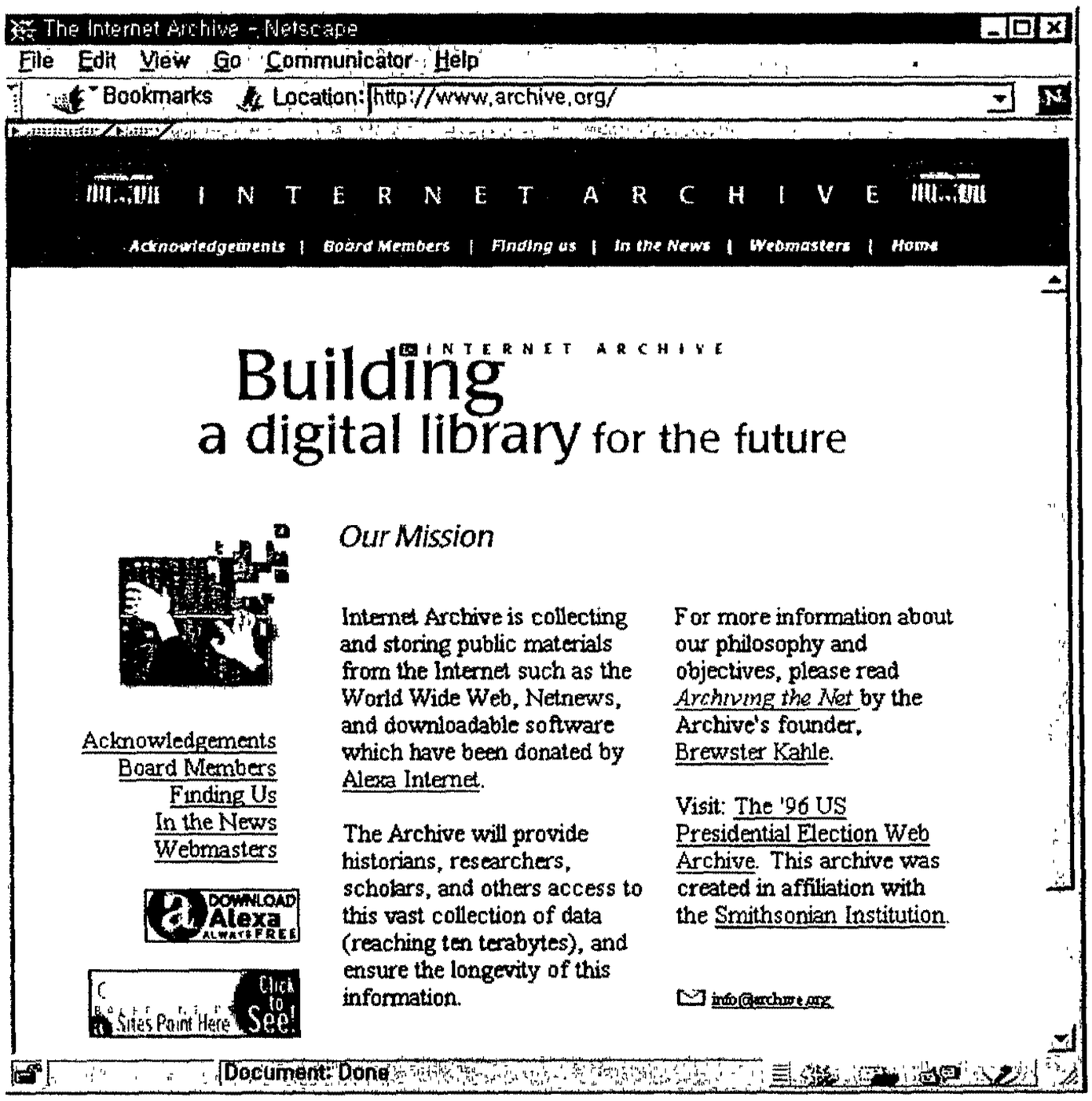
I. 서 론

사회적인 측면에서 도서관은 사회의 주요 정보자료를 수집하여 보존하고 이용시키는 역할을 담당해왔다. McGarry (1997, 254)는 이를 "사회적 기록유산의 수호자(guardian of the social memory)"로서의 역할이라고 부르면서 도서관의 핵심기능으로 보고 있다.

1990년대 들어 전자출판이 활성화되고 인터넷이 대중화되면서 사회적 기록유산의 수호자라는 도서관의 사명은 새로운 도전에 직면하게 되었다. 인쇄출판물과 달리 인터넷을 통해 제공되는 정보자료는 지속적인 이용이나 장기적인 보존이 보장되지 않는다. 내용도 수시로 갱신되므로 논문에서 인터넷 자원을 인용할 때에는 인용한 시기를 명시하도록 권장되고 있기까지 하다. 차후에 같은 제목과 내용의 자료에 다시 접근할 수 있다는 보장이 없는 것이다.

이런 문제점 때문에 인터넷 자원의 보존 및 장기적 이용에 관한 문제제기가 1990년대 초반부터 있어왔고 현재 여러 프로젝트가 진행중이다. 이중에서 특히 언론의 주목을 받은 것은 WAIS를 개발한 Brewster Kahle이 동료와 함께 1996년 4월에 설립한 Internet Archive이다.(http://www.archive.org) Internet Archive는 인터넷상의 공공자료를 수집하여 보존하고 역사가, 연구자, 학자 등에게 장기적으로 이용시키는 디지털도서관을 표방하고 있다.[그림 1] Kahle이 개인적으로 시작한 Internet Archive는 Washington Post, The New York Times, MSNBC,

〔그림 1〕 Internet Archive 홈페이지



Scientific American, Slate 등의 주요 매체에 소개되면서 널리 알려졌다. 그 결과 현재는 Sun Microsystems, Apple Computer와 같은 주요 회사와 The University of California, Berkeley의 후원을 받고 있다.

Internet Archive가 비록 대상 자원을 미국의 것만으로 표방하고 있지는 않지만 범세계적인 차원에서 기여하리라고 기대하는 것은 무리다. 실제로 Internet Archive가 스미소니안 박물관의 협조를 받아서 시험 프로젝트로 수행한 Web Archive 96 프로젝트에서는 1996년 미국 대통령 선거에 관련된 인터넷 자원만을 수집하였다.(http://www.archive.org/smithsonian.html) 프로젝트의 주체나 후원자가 모두 미국의 개인 및 단체이며 인터넷 자원의 미국 편중이 심각하다는 사실을 고려해보면 이는 어쩔 수 없는 일이다.

인터넷의 모든 자원을 수집하지 않는 한 어느 정도의 선별은 있을 수밖에 없으므로 한 국가의 인터넷 자원을 보존하고 장기적으로 이용시키는 것에 대한 책임은 결국 그 국가가 질 수밖에 없다. 실제로 영국, 스웨덴, 덴마크, 네덜란드 등의 유럽 국가와 호주 등은 자국의 인터넷 자원을 보존하는 문제를 자체적으로 해결하려고 시도하고 있다.

이 글에서는 인터넷 자원의 보존 및 장기적 이용을 위해 여러 나라에서 이루어지고 있는 국가적 차원의 주요 프로젝트를 살펴보고 우리 나라에서 시행할 모형을 모색하고자 한다.

II. 인터넷 자원의 장기적 이용을 위한 프로젝트 사례

1. 스웨덴의 Kulturarw[3] Heritage 프로젝트

1.1 배경

Kulturarw[3] 프로젝트는 스웨덴의 국가도서관인 Royal Library가 전자문헌의 장기 보존에 관해 수행하는 프로젝트로서 국가도서관 차원에서는 최초로 1996년 9월에 활동을 시작하였다.

1.2 목표

프로젝트의 목표는 출판되었다고 볼 수 있는, 온라인 상에서 접근할 수 있는 스웨덴의 전자문헌을 수집하고 보존하고 접근을 제공하기 위한 방법을 실험하는 것이다. 이 프로젝트를 통해서 현재, 그리고 앞으로의 세대를 위해서도 스웨덴의 전자 출판물의 장서의 기초를 다지게 될 것이다.

1.3 주체

스웨덴의 국가도서관인 Royal Library가 프로젝트를 이끌고 있다. 자문그룹은 National Archives of recorded Sound and Moving Images, Lund University Library, The Swedish National Archives, National Agency for Higher Education, SUNET, Sun Microsystems 등의 전문가로 구성되어 있다.

1.4 수집

인쇄된 정보는 출판사가 직접 제출하지만 인터넷 자원의 경우에는 Royal Library가 모든 자료를 능동적으로 수집하기로 결정하였다.(Arvidson, 1998)

수집대상이 되는 자료는 전자잡지와 전자신문을 비롯한 정기간행물, 정적 문헌(static document), 링크를 포함한 동

적 문헌(dynamic document)이며, 그 밖에 리스트서브에 기반한 유즈넷 그룹이나 토론그룹, FTP 아카이브, 데이터베이스도 프로젝트에 포함된다.

인터넷 로봇을 이용해 스웨덴의 인터넷 자원만을 수집하기 위해서 웹사이트의 주소가 '.se'로 끝나는 사이트를 우선 처리하고 있다. '.com'등의 다른 엔딩을 갖는 웹사이트인 경우에는 스웨덴 내의 주소나 전화번호가 있으면 스웨덴의 인터넷 자원으로 간주하여 포함한다. 이외에도 다른 나라의 다른 언어로 되어 있는 사이트에서도 스웨덴 여행을 다루거나 스웨덴 문헌의 번역물과 같은 경우는 대상에 포함시키고 있다.

일단 대상 자원으로 선택된 것에 대해서는 내용이나 품질 면에서 별도의 선정기준을 적용하지 않는다. 프로젝트에서는 로봇시스템이 모든 자료를 다운로드 받고 가능한 한 모든 것을 자동으로 하는 것을 기본으로 하고 있다. 그 이유는 다음과 같다.

- 자료의 가치를 미래의 관점에서 판단하는 것은 어려운 문제다.
- 정교한 수백만 건의 웹 페이지 선별을 위해서는 필요한 인력이 너무 많다.
- 컴퓨터 기억장치 가격이 급락했다. 보존용 로봇 시스템은 내용이 자주 바뀌는 것에 주의해서 일반적인 색인로봇에 비해 수집 주기를 짧게 잡고 있다.

1.5 메타데이터

메타데이터와 자동 편목 분야에서의 연구는 아직 준비중이다. 현 단계에서는 더블린 코어를 수용한 메타데이터 표준 개발과 자동 편목 시스템을 고려하고 있다. 자원에 대한 접근 수단으로는 URN(Uniform Resource Name) 해소 서비스를 계획하고 있다.(Arvidson, 1998)

1.6 저작권 관리

프로젝트에서 수집하여 저장한 문헌은 공개적으로 이용시키기 전에 대기시간을 둘 수 있다. Swedixh Deposit Law를 검토하기 위해 최근 구성된 위원회가 사본 보존에 있어서 저작권 문제를 고려한 권고안을 만들 예정이다.

1.7 저장

현재 데이터는 마그네틱 테이프에 저장되어 있다. 그러나, 문헌을 저장하는 방법에서 문제가 되는 것은 마그네틱 테이프나 다른 저장매체의 내구성이 아니라 소프트웨어와 하드웨어의 짧은 수명이다. 그러므로, 미래의 소프트웨어와 하드웨어 환경에서도 이용할 수 있는 장기적 저장 형태를 찾는 게 목표가 된다. 여러 포맷으로 저장하는 것도 대안의 하나로 검토하고 있다.

1.8 현황과 전망

1997년 봄과 가을로 기간을 나누어 문헌을 수집한 결과 26,000건의 웹사이트로부터 950만 건의 URL을 찾았다. 여기에는 텍스트 뿐 아니라 그림과 소리 자료도 포함된다. 스웨덴의 Kulturarw[3] 프로젝트는 다른 북유럽 국가들의 관심을 불러일으켰으며 같은 방법론을 적용한 유사한 프로젝트가 각 나라에서 시작되는 계기가 되었다. 그 결과 북유럽 국가들간의 협력을 위해 Nordic Web Archive라는 그룹이 만들어져서 정기적인 모임을 가지고 있다.

2. 덴마크의 INDOREG
(Internet DOcument REGistration)

2.1 배경

덴마크도서관센터(Danish Library Centre)는 국가 정보와 문화 유산이 인쇄물 형태로만 전달되는 것이 아니라는 인식에서 네트워크상의 출판물을 대상 자원으로 검토하게 되었다. 특히 네트워크 자원에 대한 서지제어가 인쇄물이나 CD-ROM 과 같이 물리적으로 고정된 자원에 적용되는 방식과 동일하게 이루어질 수 있는가를 밝히기 위한 프로젝트에 착수했다. 동일한 서지제어 방식을 검토한 이유는 다음과 같다.

- 현재 네트워크 상의 탐색엔진은 너무 많은 양의 데이터로 인해 부적절한 데이터를 지나치게 많이 검색해준다.
- 네트워크 자원은 물리적으로 고정된 형태의 출판물과 비교해볼 때 정보의 가치가 크게 떨어지지 않는 것도 많기 때문에 서지제어에 포함한다면 보다 많은 유용한 정보에의 접근이 동시에 이루어질 수 있다.

2.2 목표

프로젝트의 궁극적 목표는 덴마크의 국가서지라고 할 수 있는 **DanBib**에 인쇄출판물과 인터넷자원을 함께 등록하는 것이다. 국가서지의 실질적 목표를 국가적 정보 요구 만족, 자료의 수집·선택 도구, 문헌 탐색의 도구, 편목 모형으로서의 역할과 같은 네 가지로 볼 때 네트워크 자원의 등록시스템도 이 기능을 충분히 발휘하도록 마련할 예정이다.

네트워크 자원을 포함하는 국가서지 등록시스템 모형을 만들기 위해서 INDOREG 프로젝트는 수집기준, 등록방법, 유지관리, 저장이라는 네 영역에 연구의 초점을 두고 있다

2.3 주체

덴마크 국가도서관이 프로젝트의 관리자이며, 1996년 가을부터는 **National Library Authority**에서 프로젝트에 공동투자하였고 **Royal Danish Library**가 법적 보존기관과 국가서지간의 연락을 맡게 된다.(Projekt INDOREG Preface)

2.4 수집

로봇엔진을 이용하지 않고 **Danish Web Index**를 이용하여 수작업으로 네트워크상의 문헌을 수집하고 있다. 최초에 수집된 양은 60건의 홈페이지와 40건의 출판물로서 네트워크 상의 다양한 출판물 유형을 파악하기 위한 작업의 토대가 되었다.

수집되는 자료는 다음과 같이 정적 문헌과 동적 문헌으로 구분하였다.

- 정적 문헌: 정적 문헌은 내용과 서지 상세사항이 변화가 없는 문헌 유형이다. 분량으로 볼 때 도서라고 볼 수 없는 자료나 품질이 보장되지 않는 자료도 포함된다.
- 동적 문헌: 내용에 변화가 있는 문헌으로, 그 변화는 수정, 변경, 삭제, 보완 등이 모두 포함된다. 홈페이지에는 정적 문헌과 동적 문헌이 함께 포함될 수 있다.

여기에다 홈페이지 범주를 추가하여 정적 문헌, 동적 문헌, 홈페이지의 세 범주를 설정하였다. 각 범주에 적용할 선정기준은 다음과 같이 설정하였다.

① 일반적 기준: 3개 범주 모두에 적용될 기준
- 내용이 상업적, 내부적, 지역적, 사적인 것은 제외한다.
- 물리적으로 고정된 형태로도 이용 가능하다면, 네트워크상의 자료는 제외한다.
- 어느 정도의 지속성 내지 영구성을 가져야 한다. 예를 들어 페스티발 프로그램이나 광고 등은 포함시키지 않는다.

② 정적·동적 문헌을 위한 기준
- 내용 기준은 설정하지 않았다.
- 외형적 기준을 별도로 정한다.

③ 홈페이지를 위한 기준
- 내용 기준은 앞의 일반적 기준과 동일하다. 개인 홈페이지, 회사 홈페이지는 포함하지 않고 협회와 조직 홈페이지와 기관 홈페이지는 포함한다.
- 외형적 기준으로 홈페이지에 맞는 길이를 정하기는 어렵지만 적어도 6,000자 이상으로 구성되어야 하는 것으로 본다.

선정기준에 적합한 자료는 편목의 대상이 된다. 아직 편목 수준은 결정하지 않았으며 새로운 목록법을 개발하거나, 등록할 정보의 양이 매우 많을 경우 저자가 직접 메타데이터를 사용해서 직접 등록하는 방법도 검토하고 있다.

2.5 메타데이터

INDOREG 프로젝트는 출발점을 MARC 포맷에 두고서, AACR2에 따라서 편목된 MARC계열의 DanBib 데이터베이스에 인터넷 자원까지 포함하고 있다. 그러나 비물리적이고 자주 바뀌는 인터넷 자원의 특성 때문에 등록과 연관해서 여러 가지 문제가 발생하고 있다. 등록정보를 확정할 때에도 제목과 같이 그림 파일에 숨어 있거나 HTML 표준 코딩과 달라서 판단하기가 어려운 경우도 있다. 날짜 관련 필드는 주기를 첨가해서 해결하고 있다. 이러한 사정을 고려하고 인터넷 출판자의 요구를 감안하여 DanMARC2에 적절한 필드를 추가시킨 DanBib2 포맷을 만들고 이것으로 실험용 데이터베이스를 구축하였다. DanBib2에 추가된 필드는 다음과 같다.

259 File description
856 Location information for electronic material
501 Note about system requirements
512 Note on description
520 Note about bibliographic history of the work
539 Note about basis of cataloguing

NORDINFO의 후원 아래 Nordic Metadata Project에서 더블린 코어에 기반한 메타데이터 연구 프로젝트를 수행하고 있다. INDOREG 프로젝트에서는 이를 덴마크에 맞도록 변환하여 장차 네트워크 출판자가 출판물을 직접 등록할 수 있도록 할 계획이다.

ISBN 시스템에의 등록 경험에서도 보여주었듯이 전자출판물 등록 체제는 출판사들이 자연스럽게 받아들이고 있다. 자료의 등록과 갱신된 주소 정보를 유지하기 위해 덴마크 국가도서관에서는 INDOREG프로젝트를 위해 PURL (Persistent URL) 서버를 구축하였다. 출판자가 메타데이터 폼을 완성하면 자료

를 PURL 서버에 자동으로 등록시켜 국가서지 체계를 유지할 계획이다.

3. 핀란드의 EVA (Electronic Virtual Archive)

3.1 배경

EVA 프로젝트는 핀란드 교육부가 주관하는 정보사회전략 프로그램(Education, Training & Research in the Information Society)의 일부분으로서 헬싱키대학 도서관과 다른 여러 기관들의 협력 프로젝트이다. 더블린코어 메타데이터 템플릿과 변환기, URN 생성기, 수집·색인 어플리케이션 NWA(Nordic Web Archive) 등을 EVA에서 실험하게 된다. 1997년 1월에 시작되어 18개월의 기간의 일정을 가지고 있다.(Helsinki University Library. 1997)

3.2 목표

프로젝트의 목표는 다음과 같다.(Helsinki University Library. 1997)

- 전자문헌 선택 기준 강화
- 출판사로부터 전자출판물을 직접 획득하기 위한 채널, 메커니즘 강화
- 핀란드의 웹서버에서 회색문헌을 수집하기 위한 메커니즘 개발
- Finmarc, Dublin Core, 기타 메타데이터 포맷을 이용한 서지 기술 구축
- 수집한 문헌의 아카이브 구축 및 유지, 이를 위한 소프트웨어 개발
- 수집한 문헌의 출처 검증 및 탐색을 위한 구조적 메커니즘 강화
- 장기 보존 수단 연구
- 법적 아카이브 규격과 전자자료를 고려한 저작권 기준 개발

3.3 주체

참여기관과 각각의 역할은 다음과 같다.

- Helsinki University Library : 프로젝트 조정
- Jyväskylä University Library : 전자형태 보급 실험
- Oulu University Library : 전자형태 보급 실험
- Jyväskylä 대학의 Master's program in Digital Media : 문헌변환과 디지털보존
- 핀란드의 주요 출판사인 The Edita Group : 메타메이터를 가진 문헌 제공
- CSC Science Computing : 수집 기술 개발

3.4 수집방안

특별한 선정기준은 없다. 무료로 이용할 수 있는 출판된, 정형의 HTML문헌과 그림·비디오·오디오클립·애플릿 등의 인라인(inline) 자료와 뉴스리스트도 포함한다. 핀란드에서 출판된 문헌만이 수집대상이므로 네트워크 주소가 '.fi'로 끝나는 웹사이트만을 포함했다. 스웨덴의 Kulturarw[3]'프로젝트와 마찬가지로 '.fi'이외의 주소를 가지고 있는 핀란드에 위치한 웹서버들까지 포함할 수 있도록 InterNIC에 스웨덴내의 주소를 부여받은 DNS 엔트리 목록을 요청하였다.

문헌 수집은 harvester 로봇 프로그램을 사용한다. 문헌을 일단 수집한 후에는 내부에 포함된 다른 자료와 참조 링크를 분석하게 되는데, 그 위치에 상관없이 모든 멀티미디어 자료를 수집한다. 참조 링크의 경우는 그 링크가 현재

문헌의 일부가 아니라면 핀란드 사이트에 위치한 경우만 처리한다.

모든 전자문헌을 자동으로 수집할 수는 없으며 가까운 장래에 보존소법(the Act of Legal Deposit)이나 국가도서관과 출판사간의 임시 협정에 의해서 온라인 문헌의 보존소를 국가차원에서 설치하게 된다. 이 법적 보존소는 국가도서관에 설치되어 출판사들이 보내는 문헌을 보존하는 책임을 지며, 표준 파일 포맷 목록을 유지하면서 여타 포맷의 자료는 변환해서 저장할 책임도 지게 된다.

3.5 메타데이터

EVA 프로젝트는 여러 프로젝트의 집합체로서, 다음과 같이 메타데이터에 관한 다양한 도구를 시험하고 있다.

- 메타데이터 템플릿 : 저자가 직접 메타데이터를 만들 수 있도록 사용하기 쉽고 빠른 템플릿의 필요성을 일찍 인식하고 더블린코어 레코드를 생성하는 템플릿을 개발중이다.
- URN 생성기 : 국가서지번호에 기반한 URN을 만들 수 있는 URN 생성기를 고려하고 있다. 만들어진 URN 구문은 IETF URN WG이 관리하고 번호는 핀란드와 스웨덴 국가도서관이 부여한다.
- DC → MARC 변환기 : 더블린코어 레코드를 국가서지데이터베이스와 도서관 OPAC에서 함께 유지 활용할 수 있다는 것이 중요하므로 더블린코어 레코드를 MARC로 바꿀 수 있는 변환기를 실험하고 있다.

3.6 저장

초기에는 문헌과 포함자료를 따로 저장하기로 결정하였으나 나중에 함께 저장할 수 있는 적절한 파일포맷을 사용하기로 변경하였다. 수집한 문헌은 TAR와 ZIP 방식으로 압축하고 각 문헌마다 고유의 문헌번호를 부여한다.

3.7 현황과 전망

1998년 3월 현재 7,500개 웹사이트에서 약 180만 건의 문헌을 수집하였다. 86%가 일반문헌이며, 매일 5,000건의 문헌이 수집되는데 중복되는 것은 12% 정도가 되었다. 특별한 선택정책이 없이 시작하였으나, 장차 늘어나는 문헌의 크기와 보존소의 규모 증가를 고려하여 앞으로 보다 엄격한 선별규칙을 적용할 예정이다. Finmeta라는 데이터베이스를 만들기 시작했으며, 현재 더블린코어로 기술한 메타데이터 레코드가 1,000개에 이른다. 보존된 자료에 대한 접근방법은 NEDLIB 프로젝트에서 나중에 연구될 것이다.

4. 유럽연합의 NEDLIB
(Networked European Deposit Library)

4.1 배경

NEDLIB 프로젝트는 유럽연합의 응용연구프로그램 (Applied Research Programme of the European Commission)의 자금지원을 받아 유럽의 국가도서관 협력으로 수행되는 프로젝트이다.

NEDLIB프로젝트는 1998년 1월에 시작되어 36개월 동안(2000년 12월 31일까지) 진행될 예정이다. 1998년 9월까지는 일반 구조에 해당하는 전자출판물 보존체계의 필요사항을 파악하고 그 관계를 연결시킬 것이다. 정보기술전문가와 출판사 등과 활발한 토론을 도모하는 동시에 NEDLIB-INT라는 토론용 메일링리스트를 운영하여 현장으로부터의

피드백을 얻을 수 있도록 하고 있다.

4.2 목표

NEDLIB 프로젝트의 핵심 목표는 네트워크로 연결된 유럽 보존도서관(European deposit library)을 세울 수 있는 기본 하부구조를 모델링하고 검증하는 것이다. 주요 수행 과제는 다음과 같다.

- 전자출판물의 보존시스템을 위한 기능 명세
- 하부구조 모형
- 보존 지침(deposit guideline)
- 보존 시스템
- 프로젝트 보고서

4.3 주체

네덜란드 국가도서관인 Koninklijke Bibliotheek가 주도하며, 보존도서관, 아카이브, 정보기술개발자 등이 참여한다. 각 나라의 주요 참여기관은 다음과 같다.

- 네덜란드 - Agentschap ijksarchiefdienst
- 프랑스 - Bibliotheque Nationale de France
- 노르웨이 - Nasjonalbiblioteket Rana
- 핀란드 - Helsinki University Library
- 독일 - Die Deutsche Bibliothek와 CSC Ploenzke AG
- 포르투갈 - Biblioteca Nacional do Portugal와 Instituto De Engenharia de Sistemas e Computadores, INESC
- 스위스 - National Library of Switzerland
- 이탈리아 - Biblioteca Nazionale Centrale di Firenze

여기에 스폰서로 네덜란드의 Kluwer Academic과 Elsevier Science BV, 그리고 독일의 Springer-Verlag가 참여하고 있다.

4.4 현황과 전망

NEDLIB 프로젝트에서 만들어지는 일반구조모형은 기존의 국립도서관과 아카이브의 지역하부구조를 통합한 것으로 인스톨시스템, 접근시스템, 장기보존시스템의 3개의 모듈로 구성되어 실험을 거치게 된다. 수집, 저장, 메타데이터 등의 구체적인 방안은 개별 국가에서 이루어지는 프로젝트의 성과를 검토하여 결정할 계획이다.

5. 호주의 PANDORA 프로젝트

5.1 배경

PANDORA (Preserving and Accessing Networked Documentary Resource of Australia)는 호주 인터넷 출판물의 전자아카이브를 구축하려는 프로젝트로서 호주국가도서관(National Library of Australia)에 의해 1997년에 본격적으로 시작되었다. 호주국가도서관은 호주의 출판물을 수집·보존하는 법률상의 의무를 가지고 있으며, 전자출판물에도 그 논리를 확장하여 동 프로젝트의 첫 단계를 수행하고 있다.

5.2 목표

이 프로젝트의 목표는 두 가지이다. 인터넷상에서 출판되는 호주의 정보를 수집하고 보존하여 장기간의 접근을 제공하기 위한 정책과 절차를 개발하고 시험하는 것이 첫 번째 목표이며, 두 번째 목표는 전자출판물의 국가장서 구축 비용과 그 성공가능성을 검토하는 것이다. 이를 위해서 다음과 같은 개발 단계

를 계획하고 있다.

- 인터넷 문헌 탐색 및 선정
- 제작자와 연락 - 제작자의 타이틀 보존 허락을 받고 갱신주기 타이틀 형태에 대한 부수적 정보 획득
- 자료를 편목해서 국가서지데이터베이스 내로 통합, 아카이브의 엔트리 화면과 목록레코드의 PURL과 연결 제공
- 자료 복사 - HARVEST 소프트웨어 수정판을 사용, 각 자료의 엔트리화면을 만들면서 보존소 내의 개별 문헌에의 접근 제공.

5.3 선택과 접근

PANDORA 프로젝트의 진행 원칙은 선별성과 접근성의 큰 두 줄기를 이룬다. 프로젝트 초기부터 선별성의 원칙은 PANDORA선정지침의 기본을 이루었다. 프로젝트를 시작 전에 호주국가도서관은 SCOAP (Selection Committee On-line Ausralian Publications)를 설치하고 자료 선택을 위한 지침을 개발했다.(Selection Committee on Online Australian Publications. 1998-7-9) 또한 Australian Electronic Unit을 설립해서 SCOAP의 권고안을 수행하면서 중요한 타이틀을 식별, 선정, 편목하고 있다. 보존을 위해 수집되는 자료는 호주의 출판물로써 타이틀 단위로 호주국가도서관이 능동적으로 수집하는 것이다. PANDORA 아카이브에 이용자가 접근할 수 있는 유일한 접근점은 호주국가도서관의 온라인목록(OPAC)으로서 각 타이틀 엔트리화면에서는 PANDORA 표목, 제목, 출판사, 저작권사항을 제시하며, 이밖에 안내버튼, 검색버튼, 출판사 사이트와의 링크를 제공하게 된다.

5.4 저작권

법적 보존소를 다루고 있는 호주연방 저작권법은 현재 전자출판물에 관한 사항은 포함하고 있지 않다. 전자출판물을 포함하도록 개정할 것을 검토 중이므로 장차 전자출판물의 법적 보존 문제는 해결될 것이다. 저작권 문제는 멀티미디어 온라인 환경에서는 복잡한 문제이다. 동 프로젝트에서는 각 자료의 입력 화면상에서 일반 저작권 경고를 하는 동시에 출판사 측의 저작권 명시문에 연결하는 방식으로 접근하고 있다.

5.5 표준과 관리

PANDORA 프로젝트는 특별히 자원검색, 보존과 인터넷 출판, 영구적인 식별기호 부여를 위한 메타데이터표준에 관심을 기울이고 있다. 영구적인 식별기호로는 잠정적으로 유일한 해결책을 OCLC의 PURL로 판단하고 OCLC의 소프트웨어를 이용한 PURL Resolver 서비스를 설치하였다. 식별단계에서 만들어진 각 자료의 메타데이터는 호주국가도서관의 통합된 도서관자동화시스템과 국가서지데이터베이스 상에서 이용자가 접근할 수 있는 목록레코드의 기본을 이루게 된다. 호주국가도서관의 장서관리에 관한 각종 업무는 도서관의 통합자동화 시스템에 의해서 관리된다.

5.6 현황과 전망

1997년 6월 현재 180종의 자료를 선정하여 소규모의 시험 아카이브를 구축하고, 한정된 이용자에게 시험 서비스를 하고 있다. 다음 단계는 PANDORA 보존소 관리 기능을 지원하고 검토하는데 필요한 기술적 하부구조를 완성하고, 메타데이터 서버와 PURL 서버를 지

원할 프로토타입의 구현 등이 될 것이다.

이밖에도 프로젝트의 실현을 위해서는 디지털 객체 수집, 디지털 객체 관리 시스템, 보안관리 분야의 소프트웨어 개발이 필요하다. PANDORA에서 보안을 중요시하는 이유는, 호주의 문화유산에 대한 기록을 제공하는 프로젝트의 목적상 정확한 레코드가 필요하므로 보관되어 있는 문헌들은 변경이 없도록 원판 그대로 유지되어야 하기 때문이다.

6. 영국의 CEDARS

6.1 배경

CEDARS(CURL Exemplars in Digital ARchiveS)는 영국의 디지털보존프로젝트이다. 현재 영국에서는 Electronic Libraries (eLib) Programme이 진행되고 있다. 이 프로젝트의 성공을 위해서는 디지털 보존이 필요하다고 판단되었으나 현재로서는 디지털 자원의 장기 보존을 확실히 하는 공식적 메커니즘이나 디지털 출판물의 공식적 법적 보존소가 없다. 따라서 eLib 프로그램의 3번째 단계의 일부분으로 CEDARS 프로젝트가 포함되었다. 1998년 4월에 시작해서 3년의 기간동안 수행될 예정으로서 아직 계획 수준에 머물고 있다.

6.2 목표

프로젝트의 목표는 전통적으로 도서관장서에 포함되는 디지털 정보 자원을 장기 보존하기 위한 전략을 연구하며 전략적·방법론적·실제적인 문제를 제기하고 디지털보존을 위해 도서관을 위한 최상의 지침을 제공하는 것이다. 이 목표를 달성하기 위해서 프로젝트에서는 디지털 보존의 중요성을 인식시키고 전략적 프레임워크를 수립·문헌화·보급하고 디지털자원을 장기 보존하기 위한 적절한 기법을 연구할 계획이다. CEDARS 프로젝트에서는 가능한 한 eLib, JISC의 지원을 받는 다른 프로젝트를 최대한 이용하려고 한다.

6.3 주체

CEDARS는 이름 그대로, JISC(the Joint Information Systems Committee)의 자금지원을 받은 CURL 회원도서관들의 연구 계획이다. Cambridge대학, Leeds 대학, Oxford 대학이 주도하고 있으며 각각 저작권 관리와 전자잡지 대규모 온라인 DB, 메타데이터와 디지털화한 1차 자원, 보존전략으로서의 에뮬레이션의 이용과 CD-ROM 등의 자료에 관해 분담하고 있다. UKOLN은 메타데이터와 관련된 프로젝트의 일부에 참여하고 있다. 협력기관으로는 Art and Humanities Data Service, British Library, Data Archive, NPO, Research Libraries Group등이 참여하고 있다.

6.4 수집

선택기준으로는 DESIRE 프로젝트의 결과로 만들어진 Internet subject gateway를 위한 선택 기준을 원용하고 있다. 수집되는 디지털 자원의 종류는 디지털화된 1차 자원, 데이터집합, 전자잡지, 온라인데이터베이스, 프리프린트 웹페이지, 주제 게이트웨이, CD-ROM등 지적콘텐트가 구조·형태·동작에 종속되는 디지털 자원이다. 소리나 영상 형태의 정보는 고려하지 않으며, 물리적 객체 자체보다는 지적인 내용을 보존하는 것에 중점을 두고 있다.

6.5 메타데이터

CEDARS 프로젝트는 초기부터 메타데이터문제의 중요성을 인식하고 이와 관련된 문제들을 다루고 있다. 메타데이터 관련문제들을 에뮬레이션과 전환, 인증, 자원 검색, 저작권 관리, 자원평가, 관리로 나누어서 살펴볼 계획이지만 일단은 다른 연구나 프로젝트 결과를 지켜보고 있다. CEDARS내에서도 기존의 ISBN과 같은 식별기호는 계속 사용할 예정이다. 디지털 자원의 인증을 위해서 디지털 보존소 내의 각 자원에 대해 검증키값이나 체크섬 생성 같은 간단한 암호기법을 사용하는 것도 검토하고 있다.

III. 사례 분석 및 국내에서의 추진 방안

앞에서 살펴본 각 프로젝트의 주체, 시기, 수집방법, 수집범위 등을 표로 만들어 보면 [표 1]과 같다.

〔표 1〕 주요 인터넷 자원 보존 프로젝트의 비교

프로젝트	국 가	주 체	시 기	수집방법	수집범위
Kulturarw[3]	스웨덴	왕립도서관	1996.9-	로봇을 이용한 자동 수집	스웨덴의 모든 웹사이트
	전자출판물 장서의 기초. 능동적 자료 수집				
INDOREG	덴마크	덴마크도서관센터	1996-	수집. 후에 출판사의 직접 등록으로 바뀔 계획.	몇 가지사항을 제외한 모든 웹사이트
	국가서지등록 시스템 구축. 전자문헌등록과 전자문헌 제어 중심.				
EVA	핀란드	헬싱키대학 도서관	1997.1-1998.6	로봇에 의한 자동 수집	모든 웹사이트
	문헌관리를 위한 기술적 문제 고려(파일 포맷 등)				
PANDORA	호주	국가도서관	1997-	SCOAP 지침에 따른 수집	강력한 선정기준에 따른 전자출판물
	전자 보존소 구축. 선정과 접근 중심의 프로젝트.				
NEDLIB	EU	유럽 8개국 국가도서관과 출판사	1998.1-2000.12	검토 중	검토 중
	전자출판물 보존 시스템을 위한 국가간 협력, 표준 정의				
CEDARS	영국	eLib의 일부	1998.4-2000.4	DESIRE 선정원칙에 따른 수집	선정기준에 부합한 전자출판물
	eLib의 3단계 프로젝트. eLib, JISC의 다른 프로젝트와 연결.				

1. 주체

[표 1]을 보면 대부분의 국가에서 국가서지나 공동목록을 주관하는 국가중앙도서관이 프로젝트를 이끌고 있음을 알 수 있다. 우리 나라의 경우는 국립중앙도서관, 국회도서관, 첨단학술정보센터, 연구개발정보센터 등을 프로젝트 주관 기관으로 고려할 수 있다. 정보통신부의 국가전자도서관 프로젝트에 이들 기관이 공동으로 참여한 선례를 볼 때, 특정 기관이 단독으로 추진하는 것보다는 공동 프로젝트 형태를 취하는 것도 좋을 것이다. 다만 국가전자도서관 프로젝트에서처럼 자료와 시스템을 분담하는 것이 아니라 EVA나 CEDARS 프로젝트와 마찬가지로 참여기관이 과제를 분담하여 추진해야 한다.

2. 과제 분석

인터넷 자원의 보존 및 이용을 위해서 필요한 과제 및 처리과정은 [그림 2]와 같이 정리할 수 있다.

새로운 프로젝트를 공동 추진하게 되면 이 4가지 과제별로 분담 연구가 이루어지는 것이 바람직할 것이다.

3. 수집 대상 선정

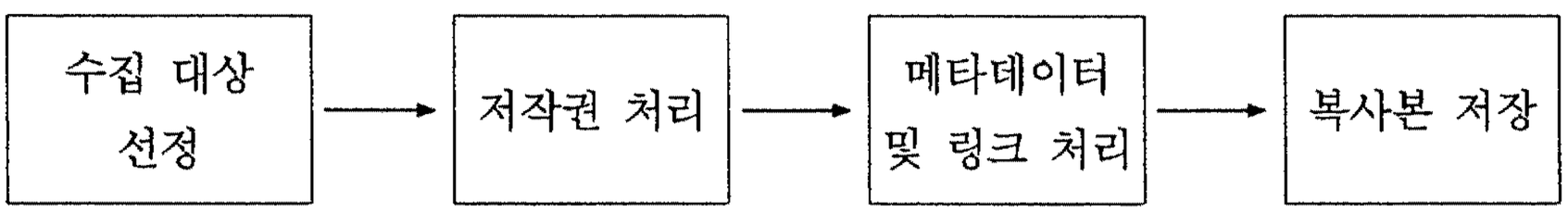
초기에는 별도의 자원 선정기준을 두지 않고 로봇엔진을 이용하여 가능한 인터넷 자원을 모두 모으는 방식이었지만, 각 프로젝트가 점차 선정원칙을 두는 방향으로 바뀌고 있다. 프로젝트를 추진할 때에는 공공성이 강한 자원부터 시작하여 점차 범위를 넓혀 가는 것도 고려할 만 하다. 최근 정보공개법의 개정, 행정정보 공동이용에 관한 규정, 한국판 GILS 프로젝트 등이 추진되고 있으므로(이춘성, 김형기, 이병만, 1998) 이와 연계하여 국내 인터넷 자원 보존 프로젝트에서 행정정보를 다룰 수도 있을 것이다.

자료의 선정은 수집 기관이 전담하는 것보다는 인터넷 검색 서비스와 마찬가지로 생산자의 직접 등록과 전담기관의 수집을 병행하는 것이 바람직하다. 인쇄출판의 경우처럼 인터넷 자원의 생산자가 메타데이터 템플릿을 통해 직접 등록하게 하는 것은 자료의 신속한 수집과 갱신이 가능하다는 장점이 있다. 반면 로봇 시스템을 이용한 수집 후 선별 방식을 통해서는 생산자가 등록하지 않더라도 널리 쓰이는 유용한 자원을 찾을 수 있다는 장점이 있다. 두 가지를 병행한다면 장점을 동시에 취할 수가 있을 것이다.

4. 저작권 처리

인터넷 자원의 납본이나 기탁을 위한 규정이 따로 마련되어 있지 않은 것은 우리 나라도 마찬가지이다. 이를 위한

〔그림 2〕 인터넷 자원의 보존 및 장기적 이용을 위한 처리 단계



법을 새로 제정하더라도 보존을 위한 자료의 등록이나 복사는 가능하게 하되, 복사본을 이용시키는 것은 원 저작자의 허락을 거치도록 하거나 일정한 시기 이후에 가능하도록 해야 한다. 이는 등록으로 인하여 원 저작자가 피해를 보지 않도록 함으로써 자발적인 등록을 유도하기 위한 필수 조건이다. 등록한 자료의 복사본은 일정 기간동안 이용시키지 않고 Nordic Web Index처럼 국가 인터넷 자원에 대한 색인을 유지하여 원 자료가 널리 이용되도록 할 필요가 있다.

5. 메타데이터 및 링크

인터넷 자원을 국가적인 차원에서 수집하기 위해서는 MARC와 같이 복잡한 형식보다는 더블린코어와 같은 간단한 메타데이터 형식이 반드시 필요하다. 생산자가 직접 등록하게 만들기 위해서는 형식이 간단해야 하며, 수집자도 복잡한 형식으로는 방대한 분량을 처리하기가 어렵기 때문이다. 인쇄출판에 비유하면 간단한 2차 정보가 포함되어 있는 판권기와 같은 역할을 메타데이터가 하게 된다. 판권기가 인쇄출판자료의 유통에 기본 데이터로 쓰이듯이 인터넷 자원의 경우에도 유통-검색 및 식별-을 위한 포맷이 필요하다. 일단 만들어진 메타데이터는 필요한 경우 수집 기관에서 좀 더 복잡한 필드를 추가하여 사용할 수도 있다. 인쇄출판물의 목록과 함께 관리할 경우에는 MARC와 매핑할 필요도 있다.

등록되는 자료는 수집 기관이 PURL이나 URN, 또는 DOI(Digital Object Identifier)와 같은 방식의 영구적인 식별기호를 부여해야 한다.

6. 복사본 저장

복사본을 저장해둘 때 고려해야 하는 것은 저장 형식과 저장 매체이다. 얼마전 한글과 컴퓨터 사의 흔글 워드프로세서 포기 사태에서 부각되었던 것처럼 특정한 저장 형식이 나중에도 계속 지원된다는 보장은 없다. 따라서 몇 개의 한정된 표준 형식으로 변환하여 저장하는 방법을 쓸 수 있다. 그러나 이런 변환에는 많은 시간과 비용이 들고 원 자료에 담긴 정보가 왜곡될 가능성이 있다. 따라서 가능하다면 원 자료를 그대로 저장하고 저장형식을 나중에라도 복원할 수 있도록 저장구조를 등록해두는 것이 바람직하다. HTML이나 SGML과 같이 표준으로 공개된 형식이 아닌 흔글이나 MS-Word와 같은 상용 프로그램에서 쓰이는 파일형식의 구조를 등록해 두기 위해서는 제작자가 이를 공개하거나 제공하도록 유도하는 별도의 조치가 필요하다.

저장 매체의 경우 영구적인 전자적 기록매체는 존재하지 않으므로 최소한 2개 이상의 매체에 저장하고 주기적인 복사를 통해 훼손을 예방해야 한다.

IV. 결 론

1998년 7월 현재 도메인 명 통계로 볼 때 앞에서 살펴본 프로젝트를 추진하는 국가들은 호스트 수가 110만이 넘는 영국이 최대이며 19만여 개인 덴마크가 최소이다.(Network Wizards, 1998) 이에 비해 우리 나라는 17만여 개 정도이므로 호스트의 수가 더 늘어나기 전

에 대책을 마련하는 것이 바람직하다.

국가적 차원에서 인터넷 자원을 보존하는 것은 자국의 자원을 발굴하여 널리 이용시키고 인터넷 자원의 서지제어를 구현할 수 있다는 측면도 중요하다. 실제로 영국의 경우 인터넷 자원의 지나친 미국 편중으로 인해 그나마 적은 자국의 정보가 잘 이용되지도 못한다고 판단하여, 주제 게이트웨이 서비스를 구축하고 자국의 인터넷 자원을 적극적으로 발굴하고 있다. (ROADS Liaison, 1998)

사회적 기록유산의 수호자로서 도서관이 계속 임무를 수행해나가려면 인터넷 자원을 제쳐놓을 수가 없으며 특히 국내 자원은 국가도서관이 나서서 자체적으로 발굴하고 보존, 이용시켜야 할 것이다.

참 고 문 헌

이춘성, 김형기, 이병만. 1998. "GILS 구현 및 메타데이터 표준화", **제5회 한국정보관리학회 학술대회 논문집**, 209-212.

Arvidson, Allan. 1998. "The Kulturarw3 Project-the Swedish Royal Web Archive", *The Electronic Library*, 16(2): 105-108.

Bennet, John C. 1997. *A Framework of Data Types and Formats, and Issues Affecting the Long Term Preservation of Digital Material.* British Library Research and Innovation Report 50, British Library Research and Innovation Centre.

Cameron, Jasmine. 1997-10-15. *Preserving & Accessing Networked DOcumentary Resources of Australia (PANDORA) Project.* National Library of Australia. <http://www.nla.gov.au/policy/plan/pandora.html>

Cameron, Jasmine, and Judith Pearce. 1998. "PANDORA at the Crossroads: Issues and Future Directions", *Proceedings of the 6th DELOS Workshop on Preservation of Digital Information.* <http://crack.inesc.pt/events/ercim/delos6/papers/pandora.doc>.

CURL Exemplars in Digital Archives Project Plan-Baseline, July 1998. 1998-8-5. <http://www.leads.ac.uk/cedars/documents/ABA02.html>.

Electronic Information Development Programme: eLib Phase 3A Digital Preservation Project Proposal: Consortium of University Research Libraries, <http://www.curl.ac.uk/Elib.pdf>.

European Commission, DG/XIII C/E. 1998-6-13. *Telematics for Libraries-Project: NEDLIB.* <http://www2.echo.lu/libraries/en/projects/nedlib.html>

Flynn, Laurie J. 1996. "Project Aims to Archive the Entire Internet", *CyberTimes*, June 14 1996.

Helsinki University Library. 1997. *EVA: the Acqusition and Archiving of Electronic Network Publications.* <http://linnea.helsinki.fi/eva/english.html>.

Internet Archive. <http://www.archive.org>.

Kahle, Brewster. 1996. "Archiving the Internet," *Scientific American*, April 96.

<http://www.archive.org/sciam_article.html>.

Kahle, Brewster. 1997. "Preserving the Internet", *Scientific American,* September 97, Special Report. <http://www.sciam.com/0397issue/0397kahle.html>.

The Kulturarw³ Heritage Project Technical Description. <http://kulturarw3.kb.se/html/projectdescription.html>.

Lounamaa, Kirsti, and Inkery Salonharju. 1998. "EVA: The Acquisition and Archiving of Electronic Network Publication In Finland", *Proceedings of the 6th DELOS Workshop on Preservation of Digital Information.* <http://crack.inesc.pt/events/ercim/delos6/papers/eva.rtf>.

McGarry, Kevin. 1997. "Library", in John Feather, and Paul Sturges (eds.), *International Encyclopedia of Information and Library Science* (London: Rougledge), 254-256.

National Library of Australia. 1998-7-9. *National Strategy for Preservation of Access to Australian Electronic Pulications: A National Library of Australia Position Paper.* <http://www.nla.gov.au/policy/paep.html>.

NEDLIB Factsheet. <http://www.konbib.nl/coop/nedlib/factsheet.html>.

NEDLIB Homepage. 1998-7-23. <http://www.konbib.ni/coop/nedlib/>.

NEDLIB News Sheet, Issue nr.1, July 1998. <http://www.konbib.nl/coop/news/newssheet1.html>.

Network Wizards. 1998. *Internet Domain Survey, July 1998.* <http://www.nw.com/zone/WWW/report.html>

Nielsen, Jorgen, 1997-9-16. *INDOREG Project Report.* <http://purl.dk/rapport/html.uk/>.

Preserving and Accessing Networked Documentary Resources of Australia Review of Progress to June 1997. 1997-10-15. <http://www.nla.gov.au/policy/pandje97.html>.

Progress Report July 1998. Cedars Document MGA03. <http://www.leads.ac.uk/cedars/documents/MGA03.html>.

ROADS Liaison. 1998. *ROADS: Resource Organisation And Discovery in Subject-based Services.* Institute for Learning and Research Technology at the University of Bristol. <http://www.ilrt.bris.ac.uk/roads/>.

Selection Committee on Online Australian Publications. 1998-7-9. *Guidelines for the Selection of Online Australian Publications Intended for Preservation by the National Library.* National Library of Australia. <http://www.nla.gov.au/1/scoap/scoapgui.html>.

Telematics Application Programme Telematics for Libraries - NEDLIB Project synopsis. <http://www2.echo.lu/libraries/en/projects/nedlib.html>