

아카이빙 데이터의 활용성 증진을 위한 전략연구: 국내외 학술논문을 중심으로*

Study on Strategies for Improving Application of Archived Data: Focused on International and Domestic Journal Articles

정영임(Youngim Jung)**

최호남(Ho-Nam Choi)***

최선희(Seon-Heui Choi)****

초 록

핵심 학술저널이 디지털화하면서 도서관에서 구독한 자료에 대한 항구 접근 및 장기 보존에 대한 요구가 증대하였다. 이러한 요구에 부응하여 국내외에서는 다양한 기관 및 기구를 수행 주체로 하여 디지털 학술자료의 보존 활동을 해오고 있다. 본 논문은 국가출연기관인 KISTI가 국내외 학술논문 아카이빙 데이터의 활용성을 증진시킬 수 있는 시스템 및 서비스 전략을 중심으로 논의하고자 한다. 또한 KISTI의 전략을 국내외 연구와 비교하였고, NDA 체제 구축의 비용편익 분석을 통해 본 연구에서 제안한 연구의 타당성을 밝히고자 하였다. 마지막으로 아카이빙 데이터 활용성 증진을 위해 정책적, 법적 기반 마련 방안과 아카이빙 데이터의 고부가가치 서비스 제공 방안을 제안함으로써 향후 연구 방향을 제시하였다.

ABSTRACT

Since core scholarly journals have been digitalized, demands on the perpetual access and long-term preservation of subscribed digital information resources by libraries are increasing. Various institutions and organizations have performed the preservation activities of digital scholarly resources for the purpose of responding those demands. This paper illustrates the National Digital Archive(NDA) system proposed and developed by KISTI and discusses on the NDA strategies which aim to improve the usability of archived journal articles. In addition, NDA strategies of KISTI have been compared with those of international researches and the economic validity of NDA has been verified by analyzing NPV, BCR and IRR. Legal system for improving the application of the archived data should be studied next and high value-added data services have been suggested as our future studies in the final section.

키워드: 항구 접근, 장기 보존, 국가 디지털 아카이브, 학술논문정보, 보존 전략
perpetual access, long-term preservation, national digital archive(NDA),
journal articles, preservation strategy

* 본 논문은 2009 한국정보관리학회 추계학술대회(2009년 11월 13일, 이화여자대학교)에서 발표한 것을 수정·보완한 것임.

** 한국과학기술정보연구원 지식기반실(acorn@kisti.re.kr) (제1저자)

*** 한국과학기술정보연구원 지식기반실(hnchoi@kisti.re.kr) (공동저자)

**** 한국과학기술정보연구원 지식기반실(sunny.choi@kisti.re.kr) (교신저자)

■ 논문접수일자: 2010년 2월 18일 ■ 최초심사일자: 2010년 2월 28일 ■ 게재확정일자: 2010년 3월 1일

■ 정보관리학회지, 27(1): 185-206, 2010. [DOI:10.3743/KOSIM.2010.27.1.185]

1. 서론

연구개발자는 연구개발 과정에서 적합한 정보의 탐색 및 분석과정에 연구기간의 30%를 투자하고, 적합한 정보를 식별하더라도 최종 목적인 전자원문 획득 경로 파악 및 획득에 어려움을 겪게 된다. 국내 학술 연구자들의 해외 연구정보에 대한 의존도가 매우 높기 때문에 전통적으로 도서관은 이러한 이용자에게는 직접적인 원문 획득을 지원하여 왔지만, 개별 도서관들이 느끼는 전자정보에 대한 라이선스 비용 부담은 증가하고 있는 상황이기 때문에 항상 자원의 선택적 수집 및 효율적 활용, 보존을 고민하여 왔다.

해외 연구정보자원은 현재 대부분 디지털화 되어 제공되고 있으나 매체의 특성상 변조가능성이 높고 출판사의 사정에 의해 사라지는 등 휘발성의 위험이 있기 때문에 이용자에게 고가의 라이선스 비용을 지불하고 획득한 자원에 대한 안정적이고 항구적인 보존 및 활용이 매우 중요해졌다. 따라서 지식 활용에 있어 중속의 심화를 방지하고 확보한 해외연구정보의 안정적·항구적·경제적 이용을 위한 국가 차원의 아카이브 구축 및 서비스 제공이 정책적·경제적·기술적으로 타당성이 있는지 검토해 볼 필요성이 있다.

한국에서는 해외출판사들의 핵심 학술지들이 전자저널화 되면서 국내 도서관들의 경제적인 전자저널 수집 및 구독을 위한 활동을 지원하기 위해 1999년 KESLI가 태동되어 활동을 하여왔고, KESLI는 그 구매력을 통하여 참여 출판사로부터 해외학술논문에 대한 메타데이터를 수집하고 이를 전자저널 라이선스 정보와 통합한

NDSL(National Science Digital Library) 서비스를 2000년 5월에 오픈하였다. 또한 2004년부터 국가 과학기술 전자정보 아카이브 구축 기본 계획(National Digital Archive, 이하 NDA라고 함)을 수립하고 만일의 사태에 대비한 전자원문 아카이빙을 시작하고 이를 활용한 서비스를 기획하기에 이르렀다. 2006년 KESLI와 NDSL 사업이 KISTI로 이관되면서 실물자료와 결합된 통합서비스인 새로운 NDSL(National Digital Science Links)로 발전하였다. 이에 기반한 KISTI의 NDA의 추진은 KESLI/NDSL과 결합하여 라이선스 보유 이용자에게는 직접적인 원문 획득을 지원하고 라이선스 비보유 이용자에게는 저렴한 비용으로 원문 획득을 지원하는 개념을 가지고 있다.

본 논문에서는 KISTI가 그동안 추진해 온 연구결과를 기반으로 확보한 국내의 학술논문 아카이빙 데이터의 활용성을 증진시킬 수 있는 시스템 및 서비스 전략을 중심으로 논의하고자 하였다. 본 연구는 학술논문의 디지털 아카이빙에 대한 관련 동향 조사를 중심으로 KISTI의 현황과 비교분석하여 향후 추진하고자 하는 사업의 전략적 방향성 및 계획 수립을 목적으로 하였다. 또한 KISTI가 확보한 데이터의 현황을 살펴보고, 이 데이터의 활용성을 확대할 수 있는 서비스 방안을 중심으로 전개하고자 하였다. 본 연구를 통해 학술논문의 디지털 아카이빙 및 이를 활용한 서비스에 대한 논의가 활성화되는 계기가 될 것으로 기대한다.

2. 선행 연구

해외에서는 디지털 아카이빙 체계 구축에 대한 다양한 측면에서의 연구가 진행되었거나 이미 진행 완료된 상태이다. 디지털 아카이빙에 대한 연구는 단순한 디지털 콘텐츠의 보존방법이나 기술에 국한되지 않으며, 디지털 아카이빙 구축 시 발생할 수 있는 다양한 문제들에 대한 분석 및 해결책 수립하고 프로토타입에 의한 검증에까지 이르고 있다. 2.1장에서는 국가별 디지털 아카이브의 동향을 분석하고, 2.2장에서는 수행 주체별 동향을 분석하였다.

2.1 국가별 디지털 아카이브의 동향 분석

2.1.1 미국

미국에서는 1996년 Task Force on Archiving of Digital Information에서 최초로 장기적인 접근이 가능한 디지털 아카이브의 국가적인 시스템에 관한 논의가 있었다. RLG(Research Libraries Group)에서는 2000년부터 2003년까지 '디지털 정보의 보존과 접근'을 중요한 사업 중의 하나로 채택하여 RLG와 DLF(Digital Library Federation)는 디지털 정보의 장기 보존을 위한 공동 위원회를 구성하였으며, 또한 RLG는 OCLC(Online Computer Library Center)와도 디지털 아카이브를 위한 하부구조 구축을 위해 협력하였다.

2.1.2 영국

영국에서는 National Preservation Office(NPO)를 설립하여 디지털 정보의 보존을 위한

국가적인 전략을 연구하고 있다. 최근의 연구에서 주관기구는 영국내의 디지털 정보의 보존에 관한 체제를 구축하고, 이러한 아카이브의 운영은 적절한 전문가를 보유한 전문기관과의 계약을 권고하고 있다. CEDARS 프로젝트는 CURL(Consortium of University Research Libraries)의 관리 및 JISC(Joint Information Systems Committee)의 자금 지원으로 1998년 4월부터 Leeds, Oxford, Cambridge 대학에서 수행하고 있다. 이 프로젝트의 목적은 디지털 컬렉션 정책의 전략 구조를 구성하고 장기 보존에 적합한 방법론을 연구하는 것이다. 이 프로젝트의 아카이브 구조는 도서관 자원이 분산된 디지털 아카이브 형태이며 OAIS 모델을 기준으로 구축되었다.

2.1.3 호주

호주에서는 국립도서관을 중심으로 표준화된 국가적 디지털 아카이브 모델을 개발하고 있다. PANDORA(Preserving and Accessing Networked Documentary Resources of Australia) 프로젝트는 국가적인 보존 차원에서 호주와 관련된 디지털 정보의 장기적인 접근을 위한 절차를 개발하고 구현하는 것이다. 특히 효율적인 방법으로 웹 사이트의 정보를 수집하고 관리하기 위한 방법론을 강조하고 있다. 이 프로젝트는 보존 시스템의 기능 구분, 보존 과정과 관련된 방법론 및 표준화의 평가, 보존비용, 장비, 인력의 추정, 디지털 보존을 위한 전략 개발, 디지털 보존을 위한 국가적인 모형의 개발을 목적으로 국립 도서관과 주립 도서관과의 협력(일부 도서관은 자신의 시스템 보유)으로 디지털 정보의 국가적인 규모의 컬렉션을 개발하고 있다. Aus-

tralian Digital Theses Project에서는 대학 도서관과의 협력을 통한 학위논문의 디지털화를 수행하고 있다. 이것은 호주 전체 도서관과의 협력을 위한 표준화된 모델이며 다른 분야의 도서관에서 다양하게 활용될 수 있을 것이다.

DSP(Digital Service Project) 프로젝트는 디지털 컬렉션을 운영하고 다른 기관들과 협력하여 디지털 컬렉션에 대한 공동의 접근을 지원하기 위한 필요조건을 구분하기 위한 것이다. 이 프로젝트는 디지털 정보의 장기 관리를 위한 하부구조를 제공하고 디지털 서비스의 효과적인 개발과 전달을 위한 기술 해결책을 도서관에서 제공하는 것을 목적으로 하고 있다.

2.1.4 일본

일본 JST는 일본 학협회와 협약 체결을 통해 자국학술지 전권의 단계적 원문 DB화(Journal@rchive)하고 있고, 해외 학술지는 국립정보학연구소(National Institute of Informatics, NII)가 디지털 아카이빙을 담당하고 있다. NII Electronic Journal Repository(NII-REO) 서비스는 Springer, Kluwer, Oxford University Press, IEEE/CS 출판사에서 제공하는 약 1,645종 저널에 대한 논문의 교차 검색 및 화면 출력을 지원하고, 도서관이나 참여 기관들이 이미 구독한 콘텐츠는 로컬 호스팅되어, 구독 저널 콘텐츠에 대해 안정적인 접근이 보장된다.

2.1.5 중국

중국과학원국가과학도서관(National Science Library, Chinese Academy of Sciences)을 중심으로 NLEng, NLM, NLA 등의 회원 도서관이 연합하여 STM 분야 학술 자원의 디지털 아

카이빙 체제를 구축하였다. 자원의 가치와 비용 분석을 바탕으로 하여 자원 수집 정책은 국가적으로 전략적 분야의 핵심 R&D 자원 가장 빈번하게 사용되고 구독되는 자원이나 희소하게 이용되더라도 중요한 자원 중 디지털 자원만을 수집하는 것을 원칙으로 한다. 오픈 소스웨어인 Fedora를 이용해 시스템 코어를 구축하고 DB는 MySQL, 검색 엔진은 Lucene, 그리고 메타데이터 표준은 Fedora METS 를 따르는 등 현 단계는 오픈 소스웨어를 이용해 구축한 프로토타입의 보존 시스템을 운영하고 있다. Nature Publishing Group, Springer, VIP의 3개 출판사로부터 1,310여 종 해외 저널을 확보하였고 중국 내 7,953종 저널을 확보하여 아카이빙하고 있다.

2.1.6 국내

국내의 경우 디지털 콘텐츠에 대한 체계적인 아카이빙은 전무한 실정이다. 현재 국립중앙도서관, 국회도서관, KISTI, KERIS 등 공공기관 및 한국학술정보, DBPIA 등의 상업기관 등에서 디지털 콘텐츠에 대한 데이터베이스화 및 관리 서비스만 진행 중이며, 장기 보존 및 이용에 대한 체계적 접근이 없는 상태로 진정한 아카이빙의 역할을 수행하지 못하고 있다(설문원 2005). <표 1>과 같이 국내 학술 저널 및 석·박사 학위논문을 주로 저장·보관하고 있어 우리나라 학술연구의 기초가 되고 있는 해외 연구정보 및 학술 논문들에 대한 진정한 의미의 디지털 아카이빙 및 보존은 이루어지지 못하고 있는 실정이다.

디지털 콘텐츠의 중요성이 인식되면서 국가적, 범국가적 차원으로 전자 학술자원에 대한 디지털 아카이빙 체제를 구축하여 운영하고 있

〈표 1〉 국내 아카이브 구축 현황

기관명	주제분야	목적	기관성격	대상컨텐츠	보존 전략	인터넷 서비스	사용자 그룹
KISTI 학회마을	과학 기술	공공	출연기관	국내 학·협회지	×	○	개인회원
국립중앙도서관	전분야	공공	국가기관	국내 웹사이트	○	○	개인회원
국가기록원	전분야	공공	국가기관	정부기록	○	○	개인회원, 국가기관
KERIS	전분야	공공	출연기관	국내 석·박사논문	×	○	대학회원
한국학술정보	전분야	상업적	민간기업	국내 학·협회지	×	○	기관회원, 개인회원
DBPIA	전분야	상업적	민간기업	국내 학·협회지	×	○	기관회원, 개인회원

다. 그러나 국내에서는 일부 디지털 자원에 대한 리포지터리를 구축하여 서비스를 하고 있으나, 디지털 자원의 영구 보존과 영속적 접근이라는 목표 아래 체계적인 아카이빙을 수행하고 있는 기관이 없는 실정이다.

2.2 수행 주체 유형별 동향 분석

2.2.1 국가 주도형

국립도서관은 기수집된 학술자원과 인쇄 학술 자원의 합법적 기탁소의 확장이라는 측면에서 국가도서관의 인프라가 디지털 아카이빙 수행에 큰 역할을 할 수 있다고 여겨진다. 특히, 상업 출판사들이 자사가 보유한 디지털 콘텐츠에 대한 장기 보존의 임무보다 출판 및 유통을 통한 수익 구조에 더욱 관심을 가지면서, 국립도서관이나 국가정보센터를 제3의 공신력 있는 기관으로 간주하고 디지털 콘텐츠 보존을 위한 파트너십을 맺고 있다.

대표적인 사례로 네덜란드 국가도서관(KB)은 소프트웨어와 하드웨어 플랫폼의 빠른 진보로 인해 위협받는 디지털 객체에 대한 장기적 접근 보장을 목적으로 2003년에 학술출판물의 디지털아카이빙 사업을 시작했다. 자발적 납본

제도를 유지하고 있는 네덜란드의 디지털 납본 자료, 웹 아카이브, 디지털화 자료 등 네덜란드 국내뿐만 아니라 전 세계 디지털출판물로 서비스 범위 확대하였으며, IBM의 디지털 보존 기술과 e-Depot의 전자출판물 아카이빙 시스템을 바탕으로 디지털 보존 체제를 성공적으로 운영해 오고 있다. 캐나다 과학기술정보연구원(Canada Institute for Scientific and Technology Information, CISTI)은 출판사와 이용권 협상을 통해 전자학술지를 개발함과 동시에 자국내 활용이 저조한 동양권 자료는 자체 수집 없이 타 기관과의 정보 협력을 통해 전자자원을 개발하고 있다. 캐나다에서 생산된 STM 발간물은 오픈엑세스 리포지터리를 구축하여 무료로 제공하고 출판사와의 협상을 통한 전자저널 원문의 보존 및 전자전송서비스(e-DDS)를 제공하고 있다. 이외에도 호주 국립도서관의 PANDORA와 일본의 NII-REO가 국가 주도형 디지털 아카이빙 사업에 포함된다.

2.2.2 개별 기관

개별 연구소, 대학 도서관 역시 디지털 자원의 보존 활동을 활발히 하고 있다. 이는 MIT에서 개발한 DSpace와 같은 개별 기관 리포지터

리 구축용 오픈 소스 소프트웨어의 개발 및 배포와 Open Archives Initiative-Protocol for Metadata Harvesting(OAI-PMH)¹⁾와 같은 오픈 프로토콜의 제정에 힘입었다.

NASA Technical Reports Server(NTRS)는 NASA의 기술문헌들을 학생, 교사, 연구자, 일반인이 이용할 수 있도록 서비스를 제공한다. 현재 NTRS에서는 약 50만 건의 항공우주 관련 서지, 온라인 원문 9만 건, 그리고 11만 건의 이미지와 비디오 등을 이용할 수 있다. DareLux(Data Archiving River Environment Luxembourg) Project는 2002년 종료된 E-Archive의 SURF project 성과에 기초하였고 이를 발전시키기 위한 프로젝트로 2004년 10월에 착수하여 2006년에 종료하였다. 이 프로젝트의 목적은 데이터셋의 장기보존을 보장할 수 있는 기술적 및 정책적 해답을 모색하는 것이다. 프로젝트의 성과로 웹 기반의 아키텍처와 OAI-PMH, XML 컨테이너를 갖춘 아카이브를 구축하였다. 이상의 기관 외에도 점차 개별 기관의 리포지터리 구축 및 자체 생산 데이터에 대한 장기 보존 활동이 활발해 지고 있으나, 모든 기관의 리포지터리가 데이터 장기 보존까지 보증할 수 있는 단계가 아니며, 개별 기관에서 수집할 수 있는 정보 자원은 한정적일 수밖에 없다. 국가적 아카이빙을 수행할 수 있는 주체라기보다는 이를 보완해 줄 수 있는 역할을 한다(Lynch 2003).

2.2.3 출판사

Elsevier, Blackwell, Highwire 등은 해외 거대 출판사를 중심으로 아카이빙 시스템을 자체 구축하여 운영하기 위한 연구를 시도하였으나, 디지털 콘텐츠의 장기 보존에 소요되는 비용 때문에 점차 디지털 아카이빙을 수행하고 있는 제3의 공신기관에 위탁하거나 각 국가 기관과 협력하는 방향으로 전환하고 있다. 이용자 입장에서 출판사들에게만 장기적 디지털 자료의 보존을 기대하는 것은 적절치 못하며 정부에 의한 국가적 수준에서 이루어져야 한다(Jones 2003).

2.2.4 제3의 공신 기관

OCLC에서 서비스 중인 Electronic Collection On-line(ECO)은 70여 출판사의 약 500여 종의 저널을 웹을 통해 서비스하고 있고, 저널의 주제 분야는 농업, 경제, 교육, 역사, 과학, 기술, 의학 등 20여 개 분야를 포함한다. 웹 아카이빙과 아카이빙을 위한 자체 시스템을 개발하여 현재 40개 출판사의 2,195종 저널 중 선택적 구독이 가능하며, 구독한 전자 저널 아카이브에 대해서는 항구 접근이 가능하다. 향후 현재의 소장 매체를 쓰지 않을 경우를 대비해 새로운 매체에 대한 갱신 및 서비스 이전에 대한 기술적 방안을 마련하고 있다.

Portico는 전자학술지의 항구적 아카이브를 제공하는 비영리 서비스로 2005년 Andrew Mellon Foundation, Ithaka, Library of Con-

1) OAI-PMH(Open Archives Initiative Protocol for Metadata Harvesting): Open Archives Initiative에서 개발된 프로토콜. 많은 아카이브들로부터 추출한 메타데이터를 사용하여 서비스들을 구축하기 위해서 어떤 아카이브에 들어있는 레코드들의 메타데이터 기술(descriptions)을 하베스트하는데 사용된다. 보통 OAI Protocol이라고 칭한다. 현재 2.0판이 나와 있는 OAI-PMH는 XML을 사용한다. Univ. of Southampton의 GNU ePrints, MIT의 DSpace 등 많은 소프트웨어시스템이 OAI-PMH를 지원한다. arXiv, CERN Document Server를 포함하여 다수의 대량 아카이브들이 이 프로토콜을 지원한다.

gree, JSTOR의 지원으로 출발하였다. 아카이빙 자료의 범위는 참여 출판사가 제공하는 학술저널로 전자형태로만 출판되는 것과 인쇄 및 전자 형태가 병행 출판되는 것, 원래 인쇄본을 디지털화한 것을 모두 포함한다. 출판사의 도산이나 출판 중지 또는 재난에 대비하고 미래에도 안정적인 접속서비스를 제공할 수 있도록 하는 것을 목적으로 한다. 출판사나 도서관에서 일정 금액을 지불하고 서비스를 받을 수 있다. 2009년 2월 Portico는 아카이브에 천만 건 이상의 저널 기사가 보존되었고, 13개 국가의 500여 도서관이 Portico의 아카이빙 서비스에 참여하고 있다.

3. KISTI NDA 연구 현황

3.1 보유한 자원 및 연구 추진 전략

KISTI는 국내 학술도서관들을 대상으로 대

규모 해외전자정보컨소시엄(KESLI)을 운영하면서 국제적 명성을 얻고 있는 유수의 국제 전자저널 파일을 제공받고 있다. 이렇게 구축된 국내의 학술지 논문들은 다양한 수록매체에 담겨져서 분산 저장되고 있으며, 장기보존전략을 위한 다각적인 방안을 고려중이다. 또한 디지털 아카이빙을 위한 전자원문을 수집하기 위하여 출판사와 라이선스 계약서를 수정하여 왔다. <표 2>는 주요 출판사의 아카이빙 관련 라이선스 계약 조건의 내용을 요약하였다. 현재까지 전자원문의 디지털 아카이빙에 동의하고 라이선스 계약을 체결한 출판사는 총 9개 출판사로서 계약내용은 모두 다소 상이하지만 원칙적으로 유사시에 보유한 전자원문을 통한 접근 및 이용을 가능하게 하고, 아카이빙 시설에 복제 및 저장까지는 허용하고 있다. 그러나 구독기간이 종료되었을 경우의 접근 및 활용 방법이나 아카이빙용 전자원문의 재배포에 대해서는 아직까지 큰 진전을 보이지는 못하고 있는

<표 2> KESLI 라이선스계약 조건 사례 분석

계약조건(1)	계약조건(2)
<ul style="list-style-type: none"> • KISTI는 A로부터 전달받은 계약학술지 콘텐츠를 KISTI에 위치한 아카이브시설에 저장 • KESLI 회원기관(구독자)은 오로지 KISTI에 위치한 도서관 아카이브시설 내의 컴퓨터터미널로부터 허가받은 이용자(학생, 교수, 직원 및 연구원)이 접근하여 이용할 수 있는 보안서버에 계약학술지 콘텐츠를 로드 및 기술적 포맷 가능 • dim archiving: 구독자와 그 허가받은 이용자들은 A가 A' 온라인서비스에 대한 접근을 3개월 이상 연속해서 제공하지 못할 경우에 그 기간에만 구독자의 보안네트워크를 통해 계약학술지콘텐츠를 접근, 탐색, 열람하고 그 일부를 인쇄 및 다운로드 가능 • 구독자와 그 허가받은 이용자들은 계약학술지콘텐츠를 실체적 또는 체계적으로 복제, 보유 혹은 재배포 불가능 • 계약학술지콘텐츠의 모든 저작권은 본 계약서에 따로 명시되지 않은 경우 A 소유 	<ul style="list-style-type: none"> • B는 저작권소유자료의 전자버전(전자저널, 전자참고자료, 전자도서 등으로서 B' System에서 이용 가능)을 인터넷으로 제공 • B는 KESLI 컨소시엄이 계약자료를 영구히 소유할 권리가 있음을 인정 • KESLI 컨소시엄은 모든 계약자료의 전자버전으로부터 전자사본 1부와 인쇄사본 1부를 백업 또는 아카이브 목적으로 복제하여 유지관리 가능. 백업 및 아카이브 목적으로 CLOCKSS 기술 사용 가능. 이러한 local archiving right는 구독자료 및 비구독자료 모두에 적용 • KESLI 컨소시엄이 구독자료 중 일부를 취소하는 경우, 허가된 이용자는 과거 구독콘텐츠를 이전과 동일한 조건으로 이용 가능 • 본 계약기간 종료로 B 출판 콘텐츠에 대한 유료접근을 중단하는 경우, B는 KESLI 컨소시엄 및 허가된 이용자에게 구독자료에 대해 구독기간 및 추가 2년 동안 출판된 부분에 대한 지속적 접근을 무료로 제공

〈표 3〉 아카이브용 원문데이터 입수현황

(기준일 : 2009.08.31)

출판사	제공범위	제공개시년	종수	2009입수건수	총누계
1	1996~현재	2006	40	354	7,300
2	1998~2006	2004	85	0	38,064
3	1997~현재	2005	19	1,020	31,474
4	1990~2005	2005	259	0	65,200
5	1995~현재	2006	2,286	413,836	4,684,673
6	창간년~현재	2006	1,288	87,952	1,182,313
7	2003~현재	2008	52	11,079	146,243
8	1997~현재	2008	65	2,034	19,917
9	1997~현재	2009	209	332,884	332,884
합계			4,303	849,159	6,508,068

형편이다.

KISTI가 보유하고 있는 아카이빙용 전자원문 데이터 현황은 〈표 3〉과 같다. 2009년 10월 말까지 총 9개 출판사의 전자원문 650만여 건을 보관하고 있다.

아카이빙의 추진을 위한 기본 원칙은 아래와 같다.

- 보존을 위한 Dark Archiving에서 활용을 위한 Light Archiving으로 이행
- 해외 주요 출판사와 라이선스 계약을 통해 해외 핵심 디지털 연구정보 확보
- 국제표준 기반의 디지털 아카이빙 시스템 구축 및 운영
- 디지털 콘텐츠 장기보존을 위한 전략적, 기술적, 경제적 연구 추진

주관기관(KISTI)이 유관기관과의 협의 하에 기본추진계획을 수립하고, 연구를 종합관리한다. 산·학·연 전문가로 구성된 기술위원회를 통해 국가 디지털 아카이브 시스템 개발에 대한 기술 자문 실시하고, 효율적인 운영을 위

하여 가칭 'NDA 운영협의회'를 구성하여 NDA의 인프라, 핵심 시스템모듈, 서비스모듈, 인터페이스 개발 관련 역할분담, 연계방안, 국제공동연구를 포함한 구체적인 개발 방안을 검토한다. 디지털 아카이브 시스템 관련 연구개발 및 서비스 경험을 지닌 해외 기관과의 국제협력체제 강화를 통해 선진 기법을 조기에 습득한다. 또한 단계별 프로토타입으로 개발된 체제를 우선 시범 서비스하고, 이를 통해 개선점을 피드백하는 선순환 체제를 구축할 예정이다. 관련 주체들의 수행 역할은 〈표 4〉와 같다.

이상에서 밝힌 기본 원칙에 따라 NDA 체제 구축 및 서비스를 위한 세부 전략은 〈표 5〉로 요약하였다.

3.2 NDA 연구 내용

해외 전자저널 연구정보의 안정적/지속적/경제적 아카이빙 및 서비스 제공을 위한 본 NDA 연구의 세부 내용은 다음과 같다.

〈표 4〉 운영주체별 수행 역할

운영주체	수행	주요 역할
NDA 운영협의회	정책 결정 및 협력기관 간 조정 전담 기구	- 기본계획 및 사업추진방침 수립 - 중점 아카이빙 분야 및 대상 확정 - 아카이빙센터의 활동 점검 및 지원
아카이빙센터 (KISTI)	해의 전자저널 아카이빙 센터 구축 운영 아카이빙 기술을 연구개발	- 예산 확보 및 출연 - 저작권 해결, 법적 제도적 장치 마련 - 협력모델 개발 및 분쟁 조절, 보존 포맷 등 표준 결정 - 기타 사업추진에 필요한 정책 수립 등 - OAIS 모델에 기반한 아카이빙시스템 구축
개별기관 (KESLI 참여기관)	NDA를 통한 개별기관 아카이빙 운영	- 데이터 수집, 제공 - 기관 내 생산 디지털 정보자원을 아카이빙센터에 제공(향후)
출판사	전자원문 제공	- 계약에 의한 전자원문의 제공 - 제공한 전자원문의 서비스 극대화를 위한 협의 및 협력

〈표 5〉 디지털 아카이빙을 위한 세부 전략

전략 범위	세부 전략
콘텐츠 수집	콘텐츠 선정 시 이용의 시급성보다 장기적 보존 및 이용성에 우선 순위를 둬
	보존 여부에 대한 결정은 수집 시 이루어져야 함
	출판사로부터 수집하는 콘텐츠의 저장 형식은 수집 시점에 필요한 보존 문제에 한정하여 결정
	개별 기관이 접수하는 전자저널 원문도 수집 대상으로 함
	콘텐츠 수집 시 저작권과 지적소유권에 대한 내용을 검토하여 권리와 의무를 명확히 함
	출판사에 대한 보상 체계 마련
	출판사에게 신뢰를 줌
	KESLI 참가기관과의 공조체제와 Buying Power 활용
콘텐츠 보존	Dark Archive 계약 시에는 한정된 용도의 이용을 통해 보관된 콘텐츠의 무결성과 이용성 점검
	아카이빙에 대한 대가의 지불 방식은 이의 활용도와 비례하도록 고안
	원문데이터와 미래의 이용을 위한 메타데이터가 함께 보존되어야 함
	저장된 데이터의 무결성은 저장일을 기준으로 주기적인 검증 실시
콘텐츠 서비스	시스템 환경에 대한 형상관리 수행
	기술의 노후화에 따른 마이그레이션의 필요성을 사전 인지할 수 있도록 이에 대한 활동을 계획에 의해 실행
	마이그레이션은 새로운 기술 환경의 표준에 대한 확신이 있을 때 추진
	서비스 대상은 KESLI 참가기관을 우선으로 하되, 미참가 기관과 개인이용자도 서비스 대상에 포함
기술	이용자 집단의 현존 및 잠재 수요를 충족할 수 있는 다양한 서비스를 개발하고 품질을 높임
	아카이빙에 대한 비용 부과는 수익자 부담 원칙에 따르며, 연간 발생 비용에 근거하여 부과
기술	서비스에 대한 평가를 통해 지속적 개선 도모
	전자저널을 비롯한 디지털 아카이빙은 많은 비용과 노력을 필요로 하는 만큼 도서관의 상호협력을 바탕으로 추진하는 것이 바람직함
기술	상호협력 체제 구축을 위해서는 아카이빙의 표준 안을 채택하고 이를 준수하는 것이 중요하며, 이를 통하여 국제적인 차원으로까지 상호협력 범위 확산 고려

〈표 5〉 디지털 아카이빙을 위한 세부 전략(계속)

기술	저장 및 보존을 위한 전자저널의 파일 형식에 대한 지침을 출판사에 제공하여 파일 형식 통제
	Preservation Description Information(메타데이터)와 Contents Information(원문데이터)는 하나의 파일에 저장하지 않고 분리하여 관리
	논문의 외부 Link 연계는 KISTI 자체적으로 구현하지 않고 출판사에서 주기적으로 보내주는 Link 업무 처리용 서버는 분산형으로, 데이터베이스 서버는 집중형으로 구성하며, 저장 장치는 Memory, Disk, Tape의 세 단계로 두어 이용 빈도에 따른 데이터 저장
	아카이브 용도의 원문 데이터는 무결성이 검증된 수집본, 보존본, 이용본의 세가지로 보관하되 기술 인력 조달 방안의 대안으로 해외의 비영리 아카이빙 기관과 제휴관계를 맺어 아카이빙 관련 기술 흡수
신뢰도 확보	해외 전자저널 아카이빙에 대한 국내의 대표적인 사업 기관으로 입지 확보
	아카이빙 책무 수행 및 운영 능력과 관련된 공식 인증 획득
홍보	출판사 홍보는 직접적인 의사 교환이 이루어 질 수 있는 수단 채택
	이용자 홍보는 잠재적 이용자에 대한 수요와 인지도 조사를 기반으로 하여 매체와 홍보 내용 결정
	정부 기관을 대상으로 하는 홍보는 아카이빙 사업 관련 정부 관계자와 여론 형성에 영향을 주는 집단으로 나누어 추진
	협력 대상이 되는 타 아카이빙 기관에의 홍보를 통해 기술 협력을 우선 도모
조직	개발 조직과 운영조직을 분리하되 초기 구축에 참여한 인력 중 일부가 운영 업무를 담당
	운영 조직 중 보존 기획(Preservation Planning) 업무 담당자는 타 업무를 겸임하지 않도록 함
재원조달	법제화에 의한 정부의 지원
	이용자에 의한 재원 조달
	정부 지원 프로젝트 수주
	KESLI Buying Power를 활용한 타 사업 수익

3.2.1 해외 핵심 디지털 연구 정보 도입

KISTI는 KESLI의 운영 주체로서 해외 대형 출판사와의 전략적 제휴를 통한 전자 저널을 비롯한 해외의 핵심 디지털 연구정보 자원을 저장·보존·이용할 수 있는 원문을 도입한다. 대형 출판사와의 원문 도입 계약을 토대로 중소규모 출판사와의 원문 도입을 가속화함으로써 〈표 6〉과 같은 규모의 해외 핵심 디지털 자원을 확보한다. 2009년 현재까지 KISTI는 Elsevier, Springer/Kluwer, BioOne, Annual Reviews 등의 대형 출판사와 원문 도입 계약을 체결하여 총 4,303종의 해외 저널의 원문 데이터를 확보하였고, IEEE, OUP, NPG 등의 출판사와도 원문 도입 체결을 위한 협상을 진행하고 있다. 또한, 해외 전자저널 뿐만 아니라

Elsevier E-book, Proquest의 미국 박사 학위 논문 확보도 추진하고 있다.

3.2.2 NDA 아카이빙 시스템 구축 및 운영

확보한 디지털 연구정보 자원이 기술 변화나 외부 위험 요인으로부터 안전하며, 장기적으로 이용될 수 있도록 아카이빙 시스템을 설계하고 운영해야 한다.

(1) NDA 기능 모델

전자저널을 비롯한 디지털 아카이빙은 많은 비용과 노력을 필요로 하는 만큼 유관 기관의 상호협력을 바탕으로 추진하는 것이 바람직하다. 상호협력 체제 구축을 위해서는 아카이빙의 표준안을 채택하고 이를 준수하는 것이 중

〈표 6〉 디지털 연구정보 자원 확보 계획

구분	2009-2010	2011-2012	2013-2014
구독 학술지 확보	1,500종	2,500종	3,600종
해외논문 호스팅	600만 건	700만 건	800만 건
해외박사 학위논문	5,000권	15,000권	25,000권
연구보고서 ²⁾	39,000권	71,000권	86,000권
디지털 콘텐츠 관리시스템	<ul style="list-style-type: none"> 출판사자료 분석 시스템 개념 설계 	<ul style="list-style-type: none"> 콘텐츠 수집관리체계 구축 콘텐츠 이용체제구축·운영 서비스 운영 개시(2011) 	<ul style="list-style-type: none"> 콘텐츠 구축 확대 부가서비스 개발 서비스 고도화 및 확산
기타	<ul style="list-style-type: none"> E-Book, 과학데이터 등의 기타정보자원에 대한 세부 추진계획 별도 수립(수요조사 및 라이선스 협약, 개발 범위 등) 		

요하며, 이를 통하여 국제적인 차원으로까지 상호협력 범위의 확산을 고려한다. 따라서 NDA의 기능 모델은 〈그림 1〉과 같이 ISO14721:2002 표준인 OAIS 참조 모델³⁾을 근간으로 하여 각 기능의 정의, Input, Process, Output, 관리 항목과 문서화 요구, 프로세스 간의 연관관계를 기술한다.

- Ingest: 정보 생산자인 출판사로부터 SIP를 접수하고, NDA 내에서 저장, 관리를 위해 콘텐츠를 검증, 변환하는 기능을 수행한다. 입수한 SIP에 대한 진본성과 품질을 확인 후, 해당 아카이브가 제시하는 포맷으로 AIP를 생성하며, 아카이브 데이터베이스에 수록할 기술정보(DI)를 AIP에서 추출한다. 또한, 수집 프로세스는 AIP를 아카이브 저장소로 전송하고, 기술정보(DI)를 데이터 관리기로 전송한다.
- Archival Storage: 아카이브 저장(Archi-

val Storage) 프로세스는 AIP의 저장, 유지 및 검색을 위한 서비스와 제 기능 수행한다. 즉, 입수한 AIP를 영구저장 공간에 추가시킨 후, 적절한 보존전략 개발, 하드웨어 에뮬레이션 실시, 보존매체 선정, 정기적인 매체 마이그레이션, 지속적인 백업과 같은 기능을 수행한다. 특히, 예기치 못한 재난을 방지하기 위하여 물리적으로 분리된 공간에 복제본을 저장하도록 하고 있으며, 아카이브 보존기간 동안 AIP의 논리적, 물리적 무결성을 유지하기 위하여 주기적으로 오류 여부를 점검하고 이용 프로세스부터의 요청이 있을 경우, 저장된 AIP의 복제본을 이용 프로세스로 전송하는 기능을 수행한다.

- Access: 이용자가 정보 요청과 접수를 할 수 있도록 하는 서비스와 기능을 수행한다. 즉, 이용자의 요청을 접수하고 이에 따라

2) 2010년 신규사업 예산을 확보할 경우의 목표로서, 매년 생산되는 과제보고서는 3만 1,000권(10년 기준)에 달하고, 연간 9.7%(과제 연평균 증가율) 증가를 감안한 달성목표임.

3) OAIS 참조 모델은 미국의 기록 보존소(NARA)는 디지털 아카이빙의 표준안을 모색하면서, 미 항공우주국의 CCSDS가 제안한 OAIS Reference Model을 표준안으로 정하였으며, 이후 많은 연구 및 사례에서 적용되고 있다(이소연 2002).

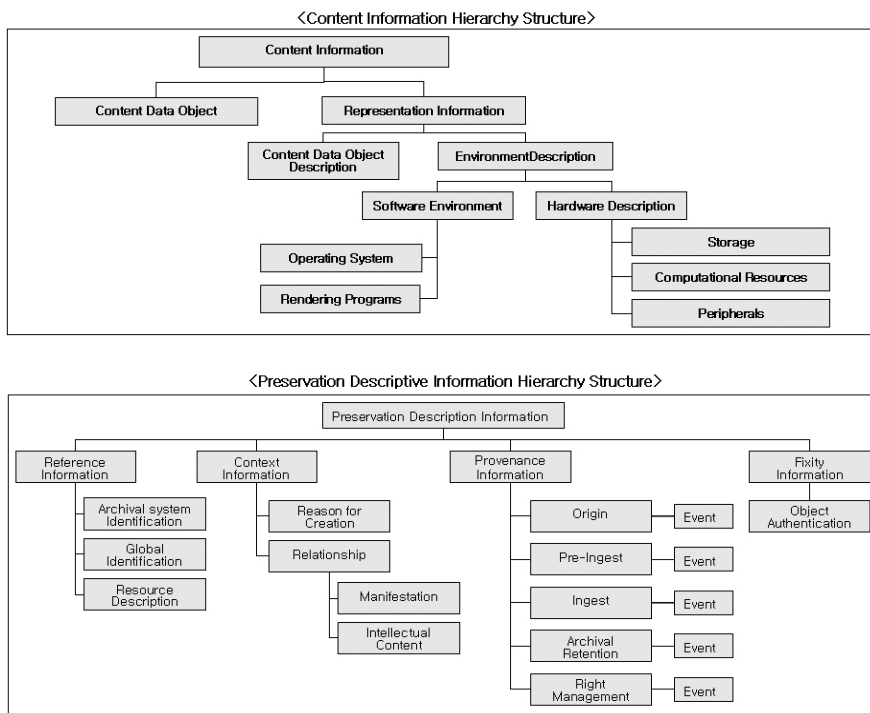
(2) NDA 데이터 모델

보존 메타데이터 프레임워크⁴⁾는 아카이브 된 디지털 객체의 보존을 위해 필요한 정보 형태를 규정하는 것으로 보존 활동에 필요한 정보 형태를 표현할 수 있어야 한다.⁵⁾

NDA 메타데이터는 <그림 2>와 같이 Content Information과 Preservation Description Information이 계층적으로 구성된다.

- Content Information: 보존대상 정보객체 (Content Data Object) 및 이의 표현 정보

(Representation Information)로 구성되며, 표현 정보(Representation Information)는 다시 환경 정보와 콘텐츠 데이터 객체에 대한 정보로 나뉜다. 콘텐츠 데이터 객체는 일련의 비트스트림 데이터이며, Representation Information은 콘텐츠 데이터 객체의 비트스트림에 대한 접근이 가능하고, 의미를 가질 수 있도록 지원하는 메타데이터이다. Representation Information은 다시 Content Data Object Description과 Envi-



<그림 2> 메타데이터 계층 구조

4) OAI Reference Model을 사용하여 아카이브를 구축한 선형 프로젝트는 CEDARS, NLA, KB 등이 있고, 최근에는 보존 메타데이터 표준인 PREMIS, METS, MODS 등을 이용하고 있다.
 5) NDA는 기존 게이트웨이 서비스인 NDSL과 통합하여 서비스를 하여야 할 것이나, 현재 운영 중인 NDSL의 메타데이터는 기본적으로 Descriptive Metadata만을 적용하고 있으므로 Preservation Metadata의 Administrative, Structural Metadata 부분이 보완되어야 한다. NDA와 NDSL 간의 통합을 위해서는 NDSL 메타데이터를 NDA의 메타데이터의 Subset으로 하여 운영하는 것이 바람직하다.

ronment Description으로 나뉜다. Content Data Object Description은 Content Data Object를 이해하고 전달하는데 필요한 Object 자체 특성, 형태를 기술하고, Environment Description은 H/W환경과 S/W 환경으로 나뉘며, Content Data Object를 표현하거나 전달하는데 필요한 기술 환경에 관한 정보를 나타낸다.

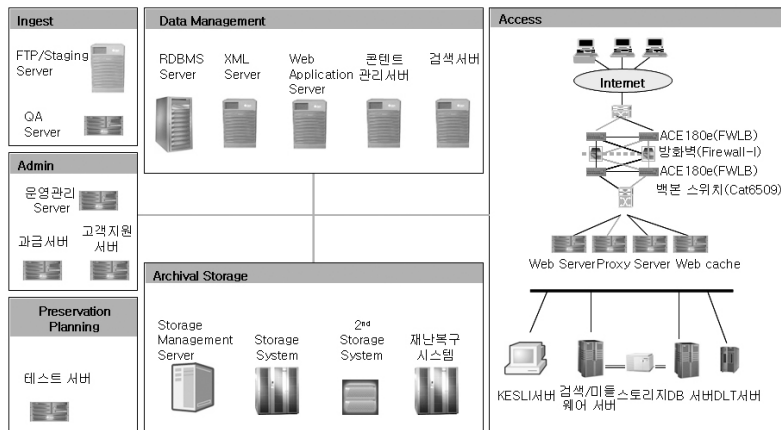
- Preservation Description Information: 아카이브된 Content Data Object의 정체성 및 관계, 유지 내역 그리고 통합관계를 기록하고 있으며, Reference, Context, Provenance 그리고 Fixity 정보로 구분된다. Reference Information은 시스템 내외부에서 고유하게 인식되는 식별자로 해당 콘텐츠를 식별하는데 필요한 정보, Context Information은 콘텐츠가 작성된 배경과 다른 정보 객체와의 연관관계를 나타내는 정보로 해당 콘텐츠에 대한 배경을 설명하는 정보, Provenance Information은 출처 및 변경 이력에 관한 정보이며, Fixity Information은 해당 객체

가 변질되지 않고 무결성을 유지하고 있는지 여부를 나타내는 정보이다.

(3) NDA 시스템 구성

NDA 시스템 구성도는 <그림 3>과 같이 OAIS의 6개 영역대로 구분하였으며, 그 개별 특성을 표현하기 위해서 각각의 시스템을 어플리케이션 특성 및 부하에 따라서 시스템의 분산화 및 통합화하여 운영한다.

- 서버: NDA의 시스템은 OAIS 모형의 Data Processing을 기준으로 하였고, 용도별로 서버를 별도로 구성한다. 운영 단계에서 각 시스템을 효율적인 어플리케이션 운영을 위하여 통합운영 또는 분리운영 등으로 재구성할 수 있다.
- 스토리지: Storage는 Ingest 과정에서 생성된 데이터를 받아 Archival Storage에 추가·저장 계층을 관리한다. 콘텐츠가 저장된 스토리지의 주기적 관리를 통하여 이용자의 요청을 만족시키는 검색과 서비스 기능 수행한다. 기본적인 저장 공간은 스토



<그림 3> NDA 시스템 구성도

리지와 백업 아카이브 시스템으로 구성되어 있으며, 스토리지 관리 시스템을 통해서 효율적인 운영을 수행한다. NDA의 시스템에서 데이터 측면을 살펴보면 가장 중요한 것이 바로 영구 보존에 대한 데이터의 안정성을 확보하는 것이다. 안정성 확보를 위해서 데이터에 대한 복제와 백업을 통한 방법이 있으며, 효과적인 데이터 보호 시스템에는 이 두 가지 방법을 병행해서 사용한다. NDA 시스템에서는 세 가지의 유형을 가지고 데이터에 대한 영구보존의 안정성을 모색한다. (1) SATA를 이용한 데이터 백업, (2) 내부 데이터 볼륨 복제, (3) 재난복구 시스템(Disaster Recovery) 활용.

- 네트워크: Front-End 영역과 Back-End 영역으로 나누어 네트워크를 구성한다. Front-End 영역은 백본 스위치를 통하여 두 개의 방화벽, L4를 통해 내부 백본으로 연결되는 구조를 가지고 있다. 향후 NDSL Front-End 영역을 공유하여 사용자가 접속하게 함으로써, 중장기적으로 네트워크의 가용성을 극대화할 필요가 있다. Back-End 영역은 Front 영역과 동일한 형태의 네트워크 구성을 취함으로써 또 다른 이중화 구현이 가능하나, 보안에 대한 대비는 성능 저하를 초래하므로 향후 서비스 제공 전략 및 목표에 따라 우선순위를 조정할 필요가 있다.

3.2.3 장기 보존 및 관리를 위한 디지털 데이터 보존 기술 연구

시간의 경과에 따른 데이터의 변형 혹은 손상으로부터 보호하기 위한 데이터 진본성 유지 및 사용 시스템 및 어플리케이션의 변경으로 인한 디지털 데이터 손실 최소화를 위한 연구를 수행한다. 보존 기술이 상당 기간 지속성을 가질 지 예측이 평가는 매우 어려운 작업이고 위험 요소를 가지고 있으므로 독자적으로 수행하기보다는 출판사 및 타 아카이빙 기관과 협력하여 수행한다.⁶⁾ 현재 가장 많이 알려진 보존 기술은 에뮬레이션(emulation) 방식과 마이그레이션(migration) 방식, 캡슐화(encapsulation) 방식의 기술로 구분된다.(Lee et. al. 2002)에 따르면 보존기술의 3가지 방식의 장단점과 디지털정보의 유형과 복잡성, 포맷, 사용 방식에 따라 적합한 보존기술을 선택하는 의사결정적인 다이어그램을 제안하고 있는데, 데이터의 적극적인 접근과 활용이 예상된다면 마이그레이션 방식이 선호된다(이수상 2004). 해외 디지털 아카이빙을 수행하고 있는 기관에서 채택한 보존 기술 및 전략은 <표 7>⁷⁾과 같다.

3.2.4 구축된 아카이빙 데이터를 활용한 부가가치 정보서비스

NDA에서 보관·관리하는 디지털 연구정보를 이용해 라이선스 만료기관 이용자들이 라이선스 기간 동안 구독했던 전자 원문에 대한 무료 이용을 보장하고 게이트웨이 서비스인 NDSL

6) 디지털 콘텐츠의 보존 전략과 관련하여 가장 활발한 연구 활동을 벌였던 네덜란드의 Digital Preservation Testbed(DPT)로 3년간의 연구 프로젝트를 수행하였다. 현재는 유럽 연합의 Planets에서 해당 연구가 수행되고 있다.

7) Kenny, A. R. et al(2006). "E-journal archiving metes and bounds: a survey of the landscape"에서 발췌하였음.

〈표 7〉 해외 기관에서 선택한 디지털 아카이빙 보존 전략

Archiving Strategy	CSI	ECO	EJC	KB	KOP	LA	LANL	NLA	OSP	PMC	PORT
Migration	○	○	○	○	○	○	○	○		○	○
Emulation				○	○			○			
Normalization	○						○	○	○	○	○
Reliance on standards	○			○	○		○	○		○	○
Refreshing	○		○	○	○			○		○	○
Use of durable media				○	○				○	○	○

을 연계하여 이용자에게 기구독 혹은 구독 중인 전자 저널에 대한 검색에서 원문 획득까지 원스탑 서비스를 제공한다. 다양한 유형의 과학기술 정보자원과의 연계를 통해 새로운 니즈에 근거한 융합 서비스 개발이 가능한데 KISTI가 보유하고 있는 연구성과, 인력, 특허, 기자재, 보고서 등의 특성화된 정보와 학술정보와의 연계를 통해 각 DB를 직접 접속하지 않고도 이용자가 검색한 논문과 관련된 다른 특성화된 정보를 찾아갈 수 있으며, 신개념의 부가가치 정보를 창출할 수 있다.

또한 개인 병원, 중소기업에 속한 연구자 등의 라이선스 비보유 이용자에게는 PPV(Pay-Per-View) 서비스를 통해 〈표 8〉과 같이 저렴한 비용으로 고급 전자학술정보의 획득을 지원하는 서비스 제공에 대해서도 출판사와 지속적으로 협상을 하고 있다.

〈표 8〉 출판사의 전자 논문 가격과 PPV 서비스 가격 비교

구분	최대(1편당)	최소(1편당)
출판사(현재)	\$40	\$20
NDA PPV(예상)	\$15	\$10

4. KISTI의 NDA 분석

4.1 KISTI의 전략을 해외동향과 비교 분석

KISTI는 NDA는 STM 분야 학술연구자를 주요 고객으로 연구정보의 공공활용을 목적으로 하는 국가 주도형 아카이브로서 제2장에서 기술한 다양한 유형 및 주체를 가진 해외 아카이브 프로젝트와 완전히 동일한 사업은 없다. 방대한 학술저널 자료의 확보나 국가 아카이브 체제 구축 및 운영은 KB와 유사하며, 데이터를 이용한 고부가가치 서비스 제공 계획은 CISTI와 유사하다. 2장에서 살펴 본 아카이빙 체계 구축 추진 유형과 NDA의 비교는 〈표 9〉와 같다.

4.2 경제성 분석

정보화 분야의 투자 사업은 경제적 편익이 직접적이고 가시적으로 발생하지 않아 경제적 타당성 분석에 어려움이 많지만 일반적으로 비용편익분석 방법에 의해 타당성을 분석하고 있다(정혜경 2004). 비용·편익분석 방법은 사업의 비용과 편익을 사회적 입장에서 측정하고 이에 따라 경제적 수익을 계산하여 사업의 타당성

〈표 9〉 디지털 아카이빙 연구 및 추진 유형

기관명	추진 주체별				목적별		대상 콘텐츠별			사용자		운영현황
	국가 주도	개별 기관	출판사	3rd party	공공 이익	상업적	정부 기록물	전자 출판물	문화 유산	자국 이용자	해외 이용자	운영 중
NDA	○				○			○		○		○
KB	○				○		○	○	○	○		○
NLA	○				○		○	○	○	○		○
CISTI	○				○		○	○		○		○
NII-REO	○				○			○		○		○
NTRS		○			○		○	○		○	○	
DareLux		○			○			○		○		
Blackwell			○			○		○			○(회원)	
Elsevier			○			○		○			○(회원)	
OCLC				○	○			○		○		○
PORTICO				○	○			○			○(회원)	○

여부를 결정하는 방법으로서 순현재가치(NPV), 비용편익비율(B/C율), 내부수익률(IRR) 등을 도출함으로써 경제적 타당성을 검토한다. NDA 사업의 비용은 콘텐츠 구축비, 장비구입비, 시스템개발비, 구축비, 유지보수비 등이 포함된다. 편익은 〈표 10〉과 같이 “정보접근성 향상, 중복

투자 방지, 연구효율성 향상, 지식정보자원 안정적 보전, 경제적 서비스제공”을 편익항목으로 선정하였다. 특정사업의 편익항목은 사업의 성격과 내용에 따라 다양하므로 사업계획에 따라 편익항목을 구체적으로 세분화하는 작업이 필요하며, 〈표 11〉과 같이 식별된 각 편익 항목

〈표 10〉 NDA 시스템 편익항목 및 측정지표

구분	편익항목	측정지표
내부	정보접근성 향상	정보접근 시간 및 장벽 제거 비용
	중복투자 방지	개별기관 아카이빙 시스템 구축 비용
	지식정보자원 안정적 보전	입수한 원문의 아카이빙 비용
	연구효율성 향상	연구 생산성 증대액
고객공통	경제적 서비스제공	해외출판사 직접 서비스 지불 가격 증대액

〈표 11〉 측정지표별 산출식

측정지표	산출식
정보접근 시간 및 장벽 제거 비용	건당 절감 시간 * 연간 이용건수 * 0.1 * 시간당 인건비
개별기관 아카이빙 시스템 구축비용	개별 기관 아카이빙 시스템 구축비용 * KESLI 참가 기관수
입수한 원문의 아카이빙 비용	사업기간내 전자원문 보존을 위한 스토리지 비용
연구생산성 증대액	연구원1인당 생산액 * 연구관리업무시간 * 일일목표생산성증가율 * 연구원 수
해외출판사 직접서비스 지불가격 증대액	(해외출판사 판매용 논문 단가 - NDA 판매용 논문 단가) * 연간 이용건수

별로 단위당 가치를 계산하여 편익을 추정하여야 한다. NDA 사업의 경제적 타당성 분석의 기간은 목표시스템 개발기간(2009-2013년)과 생명기간(7년 전후)을 포함하여 5년으로 설정하였다.

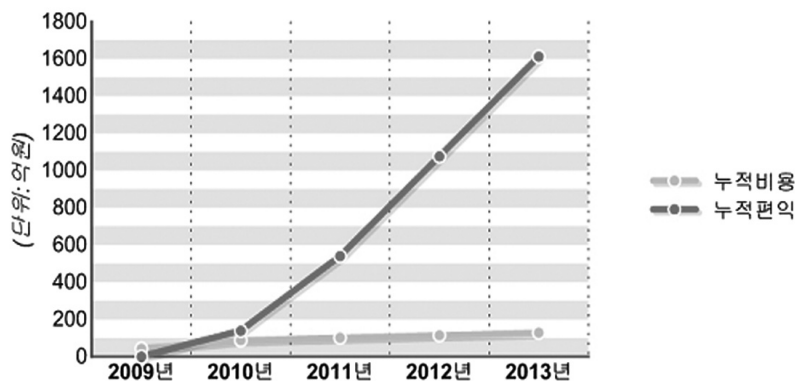
편익의 측정은 학습효과를 반영하여 차년도 의 편익 달성에 50% 영향을 주고, 차차년도 편익에 50%의 영향을 주는 것으로 가정하였다. 6차년도(2014년) 이후는 전체 시스템이 완성 되어 2년이 경과한 시기이므로 모든 시스템에서 편익은 100% 발생하는 것으로 가정하였다. 비용은 사업 초기년도부터 2차년도까지 집중적

으로 발생하며, 이후 2013년까지 시스템 유지 보수 비용이 발생하게 된다. 편익은 사업 개시 후 2차년도(2010년)부터 발생하기 시작하여 이후 2012년까지 큰 폭으로 증가할 것으로 전망되며, 이후 완만하게 증가할 것으로 추정하였다. 전체 사업기간 중 발생하는 비용과 편익의 추세는 <그림 4>와 같이 나타났다. 사업개시 후 2차년도(2010년)부터 누적편익이 누적비용을 초과하여 순편익이(+)로 전환되며, <표 12>와 같이 이후 그 폭이 점차 확대될 것으로 추정 된다.

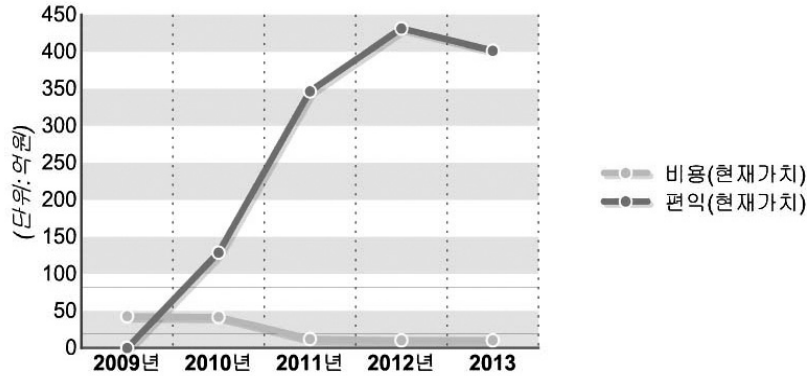
<표 12> 추정 비용·편익 비교(할인율 7.5% 적용시)

(단위: 억 원)

	2009년	2010년	2011년	2012년	2013년	합계
비 용	43	45	14	13	14	129
편 익	0	139	401	536	536	1611
누적비용	43	88	102	115	129	-
누적편익	0	139	540	1076	1612	-
순편익	-43	94	387	523	522	1482
비용(현재가치)	43	41.9	12.1	10.5	10.5	117.9
편익(현재가치)	0	128.8	346.6	431.5	401.4	1308
NPV						1,190
BCR						11.09
IRR						373%



<그림 4> 국가 디지털 아카이브 시스템 구축사업: 누적비용·누적편익



〈그림 5〉 국가 디지털 아카이브 시스템 구축사업: 비용·편익(순현재가치)

본 사업의 비용과 편익을 순현재가치로 환산하여 비교하면 〈그림 5〉와 같으며, 비용의 순현재가치는 2009년부터 2010년까지 매년 41~43억 원 범위 내에서, 2011년부터 2013년까지는 매년 10~12억 원 범위로 지속적으로 투입되며, 이에 따른 편익은 편익반영비율이 100%에 이르는 2012년까지 급격하게 증가하고 그 이후로 완만하게 감소하는 것으로 분석되었다. 또한 사업의 총 현재가치(NPV)는 기준 할인율 7.5% 적용시 1,190억 원으로 나타나고, 최대 할인율 15.0% 적용시 973억 원, 최저 할인율 5.0% 적용시 1,278억 원으로서, 973~1,278억 원의 범위로 나타났다. 사업의 편익/비용 비율(BCR)은 기준 할인율 7.5% 적용시 11.09배(1109%)로 나타나고, 최대 할인율 15.0% 적용시 9.9배(990%), 최저 할인율 적용시 11.53배(1,153%)로서, 최저 9.9배에서 최대 11.53배까지로 산출되었다.

비용편익 분석에 따른 결과를 최종적으로 종합평가해 보면, NDA 시스템 구축사업은 할인율 7.5%를 적용한 경우를 기준으로 볼 때, 다음과 같이 순현재가치(NPV), 편익/비용 비율(BCR),

내부수익률(IRR) 등 3가지 측면에서 모두 경제적 타당성이 높은 사업으로 평가되었다.

5. 결론 및 제언

본 논문에서는 기구독한 디지털 학술정보 자원의 항구 접근 및 장기 보존에 대한 요구에 부응하기 위해 국가정보센터인 KISTI를 수행 주체로 하여 국가 디지털 아카이빙 체제 구축을 위한 연구 전략과 내용을 제안하였다. 본 NDA 체제 구축의 세부 전략과 시스템은 학술정보자원의 항구 접근 및 장기 보존을 목표로 할 뿐만 아니라 아카이브된 자원의 활용성을 고려하여 고안되었다. 또한, 본 연구 내용의 독창성과 타당성을 검증하기 위해 국내외 아카이빙 연구와 비교 분석하였고, 비용·편익 분석에 따라 연구 사업의 객관적 타당성을 증명하였다.

마지막으로 아카이브된 학술정보자원의 향후 재활용성 증진을 위한 고려사항은 다음과 같다.

첫째로, 디지털 아카이빙 체제 구축을 위해

수집하고자 하는 각종 정보자원은 누군가의 지적 재산이며 따라서 이들 자원의 보존하고 활용하기 위한 일련의 행위는 다른 사람의 지적 재산권과 관련한 여러 가지 문제들을 야기할 수 있다. 따라서 저작권법과 납본 관련법, 그리고 실제 디지털 아카이빙 체제 구축을 위한 일련의 절차나 기구 등에 관한 법적 기반을 마련하기 위한 관련법 등 여러 형태의 법제를 이해하고, 필요한 경우 저작권을 소유한 기관과 협력 및 계약 체결을 통해 과학기술분야 디지털 아카이빙 체제의 정책적 기반을 마련하는 것은 해결해야 할 중요한 과제 중 하나이다.

둘째로, 디지털 아카이빙 체제의 안정적 운영에 소요되는 비용 마련을 위해서는 아카이빙 데이터를 활용한 독자적 수익 모델을 마련해야 한다. 또한 국내의 학술논문 원문을 직접 보관하게 되면 학술연구자로 대표되는 이용자 집단의 현존 및 잠재 수요를 충족할 수 있는 다양한 서비스를 개발하고 데이터의 품질을 높일 수 있어 과학기술 분야의 고품질·고부가가치의 데이터 서비스를 제공함으로써 KISTI가 “과학기술정보의 체계적 수집, 유통, 활용과 연구개발 서비스 활성화”에 이바지할 수 있을 것이다.

고부가가치 데이터는 학술논문 원문에서 다양한 정보를 추출하고 이를 이용자의 요구에 맞게 재가공함으로써 생성된다. 과학기술분야의 학술논문이나 연구보고서 원문에 나타나는

표, 그림, 차트 등의 정보는 각종 통계 및 수치, 경향, 분포 등을 집약적으로 보여주는 고품질의 데이터이다. 때때로 이용자는 학술논문 전체보다는 특정 표, 그림, 차트만을 필요로 하는 경우도 있는데, 전자 원문에서 이러한 정보를 추출하기 위해서는 전자 원문의 파일 포맷이 XML, HTML 혹은 SGML과 같이 태그가 부착된 구조화된 포맷이어야 한다.⁸⁾ 현재 HTML 문서에서 의미 있는 표를 추출하는 테이블 마이닝 기법에 대한 연구가 활발하게 이루어져 이미 높은 정확도를 보이는 기술이 등장하고 있다(Jung 2006). 키워드-문서 간 유사도를 기반으로 하는 정보 검색과는 달리 정보 추출은 문서의 해당 언어 특성에 크게 의존하기 때문에 자연언어처리 기술을 도입하여야 하며, 텍스트 내 특정 패턴을 가지는 정보는 다양한 텍스트 마이닝, 데이터 마이닝 기법을 활용하여 추출할 수 있다. 학술논문에서 고부가가치 정보 추출에 대한 연구는 캐나다 CISTI에서도 관심을 가지고 진행하고 있으며, 이를 통해 다양한 고품질의 데이터 서비스 제공이 가능할 것이라 기대하고 있다.⁹⁾ 또한, 특허 문서에 대한 영한 특허 번역 서비스처럼 영문 학술논문의 초록 및 원문의 일부를 번역해주는 서비스를 개발한다면, 보다 많은 논문의 내용을 신속하게 파악하고자 하는 국내 학술연구자에게 큰 도움이 될 것이다.

8) A출판사는 KISTI에 학술논문의 XML 포맷 원문을 제공하고 데이터 서비스에 활용할 수 있도록 하였다.

9) KISTI 국제수준진단 현장실사 심사자로 왔던 CISTI의 Dr. Bernard Dumouchel과 KESLI 담당자와의 인터뷰 내용 중 발췌하였음.

참 고 문 헌

- 설문원. 2005. 과학기술 분야 디지털 아카이빙을 위한 정책연구. 『정보관리연구』, (36)4: 23-49.
- 이소연. 2002. 디지털 아카이빙의 표준화와 OAIS 참조모형. 『정보관리연구』, (33)3: 45-68.
- 이수상. 2004. 디지털 아카이빙의 워크플로우와 보존처리 기술에 관한 연구. 『한국도서관·정보학회지』, (35)3: 119-138.
- 정혜경. 2004. 디지털 아카이빙의 경제성 분석 연구. 『한국문헌정보학회지』, (38)4: 251-270.
- 한국과학기술정보연구원. 2005. 『국가 디지털 아카이빙 체제 구축에 관한 연구』. 한국과학기술정보연구원.
- 한국과학기술정보연구원. 2007. 『해외 연구정보 국가 디지털 아카이브 구축』. 한국과학기술정보연구원.
- 한국과학기술정보연구원. 2008. 『해외 전자저널 항구보존 및 공동활용체제 구축의 타당성 분석』. 한국과학기술정보연구원.
- 한국저작권위원회. 2009. 『개정 저작권법 해설』. 문양사: 서울.
- “DareLux” [online]. [cited 2009.9.30]. <<http://www.surffoundation.nl/en/projecten/Pages/DARELUX.aspx>>.
- Digital Preservation Testbed, 2003. From Digital Volatility to Digital Permanence: preserving text documents. [online]. [cited 2009.8.5] <<http://www.digitaleduurzaamheid.nl/bibliotheek/docs/volatility-permanence-textdocs-en.pdf>>.
- Dumouchel, B. 2009. Interview by author, 14, October, KISTI. AV recording.
- Jones, M. 2003. Archiving E-Journals Consultancy-Final Report. [online]. [cited 2009.9.13] <http://www.jisc.ac.uk/uploaded_documents/ejournalsfinal.pdf>.
- Jung, S. and H.-C. Kwon. 2006. “A Scalable Hybrid Approach for Extracting Head Components from Web Tables,” *IEEE Transactions on Knowledge and Data Engineering*, 18(2): 174-187.
- Kenney, A. R., R. Entlich, P. B. Hirtle, N. Y. McGovern, and E. L. Buckley. 2006. E-Journal Archiving Metes and Bounds: A Survey of the Landscape [online]. [cited 2009.8.5]. <<http://www.clir.org/pubs/reports/pub138/pub138.pdf>>.
- Lee, K., O. Slattery, R. Lu, X. Tang, and V. McCrary. 2002. “The State of the Art and Practice in Digital preservation.” *Journal of Research of the National Institute of Standards and Technology*, 107(1): 93-106.
- Li, C., X. Zhang, and Z. Wu. 2007. “Preservation Management in Practice: Trusted Workflow.” Presentation in International Conference on Preservation of Digital

- Objects, iPress2007, Oct [online]. [cited 2009,10,15].
<<http://ipres.las.ac.cn/program.jsp>>.
- Lynch, C. 2003. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." *ARL Bimonthly Report*, 226. [online]. [cited 2009,7,2].
<<http://www.arl.org/newsltr/226/ir.html>>.
- NASA Technical Report Server 2009. "Collections in NTRS". [online]. [cited 2009, 9,30].
<<http://ntrs.larc.nasa.gov/index.jsp?method=collections>>.
- Zhang, X. 2007. "Chinese Digital Archive Network of Foreign STM Literature." Presentation in International Conference on Preservation of Digital Objects, iPress2007, Oct [online]. [cited 2009, 10,15].
<<http://ipres.las.ac.cn/program.jsp>>.