

디지털 아카이브의 영구적 보존을 위한 개념적 모형 설계에 관한 연구*

Conceptual Model for Developing Persistent Digital Archives

서 은 경(Eun-Gyoung Seo)**

목 차

- | | |
|---------------------------|---------------------------|
| 1. 서 론 | 2. 7 Knowledge Bank 아키텍처 |
| 2. 디지털 아카이브의 기능적 아키텍처 | 3. 영구적 디지털 아카이브 구축을 위한 전략 |
| 2. 1 OAIS 기능모형 | 3. 1 신뢰성 높은 디지털 아카이브 |
| 2. 2 InterPARES 보존기능 모형 | 3. 2 지속성 높은 디지털 아카이브 |
| 2. 3 IRMS의 보존기능 아키텍처 | 3. 3 보안성 높은 디지털 아카이브 |
| 2. 4 NAL 디지털 아카이브 아키텍처 | 4. 디지털 아카이브 개념적 모형 설계 |
| 2. 5 NEDLIB 디지털 아카이브 아키텍처 | 5. 결 론 |
| 2. 6 DSpace 아키텍처 | |

초 록

보다 효과적으로 디지털 아카이브를 구축하고 디지털 아카이브간의 상호운영과 자원의 공유를 수월하게 수행할 수 있도록 하는 표준적인 아카이브 모형과 이에 따른 정책 및 방법을 제안하는 OAIS 참조모형이 권고되었다. OAIS 참조 모형은 개념적으로 최상위 차원에서 정의한 것으로 현재 개발된 디지털 아카이브는 적용시키려는 모 기관의 목적, 시스템 환경 및 제반환경과 같은 다양한 요소들을 고려하여 개발되고 있다. 이에 현재 구축되어진 디지털 아카이브 아키텍처를 조사하여 앞으로 구축될 디지털 아카이브가 갖추어야 할 지녀야 할 기술적 전략을 제시하였다. 이 결과를 토대로 하여 본 연구는 OAIS 참조모형이 권고한 기본개념과 기능, 환경을 근간으로 하나 보다 신뢰성과 보안성이 높고 정보기술 관리기능이 영구적으로 유지될 수 있는 디지털 아카이브의 개념적 모형을 제시하였다.

ABSTRACT

The Reference Model for an Open Archival Information system(OAIS) was recommended for developing digital archives effectively and for implementing interoperability and resource sharing between digital archives. But as the Reference Model is high-level recommendation, it does not specify a design and an implementation. Therefore, this study investigates the architecture of several digital archives, suggests three strategies for persistent archive, and proposes the conceptual model for developing trusted, persistent, and secure digital archive.

키워드: 디지털 아카이브, 디지털 아카이브 아키텍처, OAIS 참조모형, 영구적 보존, 신뢰성 높은 아카이브, 보안성 높은 아카이브, 영구적 아카이브

Digital Archive, OAIS Reference Model, Trusted Digital Archive, Persistent Digital Archive, Secure Digital Archive

* 이 논문은 2003년도 한성대학교 교내연구비 지원과제임.

** 한성대학교 지식정보학부 교수(egseo@hansung.ac.kr)

논문접수일자 2003년 11월 20일

게재확정일자 2004년 2월 20일

1. 서론

정보기술과 인터넷의 급속한 발전과 확산은 정보환경을 다각적으로 변화시키고 있다. 이제 이용자는 인쇄매체가 아닌 전자매체에 구축된 정보자원 즉 디지털 정보자원을 손쉽게 접근하여 활용하고 있으며 그 이용도와 의존도는 점차적으로 확장되고 있는 실정이다. 디지털 정보자원이란 디지털 형태의 텍스트, 오디오, 비디오 등과 같은 디지털 객체(objects)를 가리키는 것(Witten & Bainbridge 2003)으로 그 예로 전자잡지, 전자책, 웹사이트, 디지털사운드, 디지털 동영상 그리고 각 기관에 축적한 각종 디지털 문서 등을 들 수 있다. 현재 디지털 도서관을 비롯하여 기록관, 출판사, 방송국, 정보제공기관과 같은 기관들은 디지털 정보자원의 체계적인 수집/생성, 소장/축적, 접근/검색, 전송/배포, 이용, 공유를 위한 디지털 환경을 수립하는데 그 목적을 두고 시스템 구축과 원격정보 제공에 많은 노력을 기울이고 있다. 그 반면에 이용자들이 미래에서도 그 정보자원을 계속 접근하여 활용할 수 있게 하는 보존 기능 특히 디지털 정보자원의 장기보존을 위한 기술적 처리와 시스템 구현은 상대적으로 적은 비중을 차지하고 있고 대부분의 시스템에서 간과되고 있는 실정이다.

전통적인 의미로서의 보존이란 보존대상을 형태와 콘텐츠면에서 변화시키지 않고 있는 그대로 계속 유지하는 것을 뜻하나 디지털 환경에서는 데이터의 포맷과 매체의 변환은 인정하고 콘텐츠만을 장기적으로 더 나아가서 영구적으로 유지시키는 활동을 말한다. 이는 형태적 변환 없이 있는 그대로 소장하여 보존

한다면 정보의 접근은 점점 힘들어 지거나 불가능해지기 때문이다. 특히 Hodge(2000)는 디지털 콘텐츠를 영구적으로 보존시키는 기능 뿐만 아니라 디지털 정보자원이 생성시 지닌 모양과 느낌까지도 계속해서 유지시키는 동시에 그 콘텐츠를 사용케 해주는 여러 가지 메커니즘, 즉 원문에 대한 액세스, 검색, 처리기능, 역동적이고 상호작용적인 전달기능, 이용자 기반의 원문 재생기능 등이 사라지지 않게 유지시키는 모든 작업까지 일컫는 것을 디지털 정보자원 보존이라 하였다. 따라서 디지털 보존은 생성된 정보 그대로를 보존하는 기능과 동시에 동적으로 그리고 진보된 기술로 디지털 정보를 접근할 수 있게 하는 기능에 의하여 이루어 질 수 있다(RLG/OCLC Working Group 2002, p.3).

디지털 정보자원의 장기적 보존은 정보기술적인 문제와 경제적 및 정책적인 문제를 지닌 연구영역으로 현재 다각적인 관점에서 연구가 수행되고 있으나 아직까지 뚜렷한 해답과 전략이 제시되고 있지 않는 실정이다. 다만, 장기보존 전략을 제시하는 여러 연구에서 디지털 정보자원을 효과적으로 그리고 장기적으로 보존하기 위해서는 디지털 정보자원의 생명주기별로 관리하는 디지털 아카이빙(digital archiving) 기능이 수립되어야 한다고 제안하고 있다(Brown 2000, Hodge 2000, Rossell 2000, Berthon et al. 2002).

디지털 아카이빙은 좁은 의미로 컴퓨터 파일의 일시적 백업 컬렉션이나 데이터를 정리 통합하여 처리 가능한 형태로 만든 정보파일 또는 그 집합체(정진택 2001)를 의미하는 것으로 사용되어 왔으나 디지털 보존 환경아래

에서는 지속적인 가치를 가졌다고 판단되는 디지털 객체를 장기간 관리하는 활동(이소연 2002) 또는 가치 있는 디지털 정보자원을 선별하여 디지털 콘텐츠와 여러 기능들을 디지털 정보자원의 생명주기별 보존 관리하여 미래 이용자들이 어려움 없이 정보를 활용할 수 있도록 하는 전반적인 작업(서은경 2003)으로 정의내릴 수 있다. 이러한 디지털 아카이빙 작업을 수행하는 시스템이 바로 디지털 아카이브이다(RLG/OCLC Working Group 2002, p.3). 디지털 환경에서의 아카이브는 다양한 디지털 정보자원의 소장 및 보존, 접근기능을 지닌 시스템 또는 그러한 기능을 가진 프로그램 자체를 일컫는 것으로(CCSDS 2002), 일반적으로 다양한 유형의 디지털 정보를 체계적이며 효율적으로 보존·활용할 수 있도록 하기 위하여 디지털 정보자원의 평가 및 선택, 데이터 관리 및 기술, 보존, 접근, 네트워크 기반의 시스템 운영, 시스템을 통한 생산자와 이용자의 연결을 지원하는 전반적 시스템을 지칭한다.

1996년 이후로 디지털 아카이브는 대규모 디지털 정보자원을 다루는 국가대표기관에서 다년간 국가적 프로젝트의 성과를 토대로 하여 각각 구축되어져 왔다. 최근 국가기관뿐만 아니라 각 도서관이나 디지털 정보를 축적 보존하여 정보서비스를 하고자 하는 다양한 기관에서도 각각의 목적에 적합한 디지털 아카이브를 구축하고자 하는 시도가 발생되자, 보다 효과적으로 디지털 아카이브를 구축하고 디지털 아카이브간의 상호운영과 자원의 공유를 수월하게 수행할 수 있도록 하는 표준적인 아카이브 모형과 이에 따른 정책 및 방법이 제안될

필요성이 생겨졌다. 이에 NASA(National Aeronautics and Space Administration)와 CCSDS(Consultative Committee for Space Data Systems)는 한 아카이브가 디지털 정보를 영구적이거나 무기한 장기적으로 보존하는데 필요한 기술적·정책적 권고안을 개발하여 1999년에 OAIS 참조모형(Reference Model for an Open Archival Information System)을 제안하였고 이 참조모형은 2002년에 ISO표준(1471:2002)으로 확정 공포되었다(김희정 2003).

OAIS 참조모형은 개념적으로 최상위 차원에서 디지털 아카이빙을 정의하는 것으로 앞으로 구축될 디지털 아카이브는 참조모형이 제시한 기능과 환경을 따라야 하지만, 적용시 키려는 모 기관의 목적, 시스템 환경 및 제반 환경과 같은 다양한 요소들을 고려하여 개발되어야 할 것이다. 본 연구는 디지털 아카이브는 OAIS 참조모형이 권고한 기본개념과 기능, 환경을 근간으로 개발된 시스템이어야 하나 보다 신뢰성 높고 영구적으로 유지되는 디지털 아카이브가 되어야 한다는 전제아래 영구적 디지털 보존을 수행할 수 있는 기술적 전략과 개념적 모형을 제시하였다. 이를 위하여 먼저 여러 연구에서 개발된 디지털 아카이브의 개념적 모형과 아키텍처를 살펴보고 다음, 영구적 디지털 아카이브 구축을 위하여 본 연구가 제시한 세 가지 전략을 수용하는 디지털 아카이브의 개념적 모형을 설계하였다. 본 연구가 제안된 기술적 전략 및 개념적 모형이 앞으로 우리나라에서 대규모 디지털 아카이브를 구축하는데 기초적 역할을 할 수 있을 것으로 판단된다.

2. 디지털 아카이브의 기능적 아키텍처

본 장에서는 디지털 아카이브가 갖추어야 할 기본적 기능과 시스템 아키텍처를 파악하기 위하여 OAIS와 InterPARES가 제안한 기능 모형과 현재 구축되어 운영 중이거나 개발 중인 5개의 대표적 디지털 아카이브의 사례를 살펴보았다.

2.1 OAIS 기능 모형

OAIS는 디지털 아카이브를 장기적으로 유지 및 관리하기 위해서 '흡수', '보존', '데이터관리', '운영', '보존계획', '접근' 과 같은 6 가지 기본적 기능이 유기적으로 이루어져야 한다고 제시하였다(CCSDS 2002). 각 모듈의 기능을 간단히 살펴보면 다음과 같다. 흡수모듈은 생산자로부터 유통정보패키지의 입수, 소장, 관리를 위한 전초작업을 실행한다. 즉 입수한 유통정보패키지에 대한 진본성과 품질을 확인한 다음, 해당 아카이브가 제시하는 포맷으로 보존정보패키지를 생성하며, 이때 아카이브 데이터베이스에 수록시킬 기술정보를 보존정보패키지에서 추출한다. 또한 흡수모듈은 보존정보패키지를 아카이브 보존모듈로 전송하고 기술정보를 데이터 관리모듈로 전송시킨다. 아카이브 보존모듈은 보존정보패키지의 저장, 유지, 그리고 검색을 위한 서비스와 제 기능을 수행한다. 즉 입수한 보존정보패키지를 영구저장 공간에 추가시킨 다음, 적절한 보존전략 개발, 하드웨어 에뮬레이션 실시, 적절한 보존매체 선정, 정기적 매체이전, 지속적인

백업과 같은 기능을 수행한다. 특히, 예기치 못한 재난을 방지하기 위하여 물리적으로 분리된 공간에 복제본을 저장하도록 하고 있으며, 아카이브 보존기간 동안 보존정보패키지의 논리적·물리적 무결성을 유지하기 위하여 주기적으로 오류 여부를 점검하고 접근모듈로부터의 요청이 있을 경우, 저장된 보존정보패키지의 복제본을 접근모듈로 전송하는 기능을 담당한다.

데이터 관리모듈은 보존정보패키지의 기술정보를 확인하고 아카이브를 운영하는데 필요한 행정정보를 유지하는 기능을 한다. 즉 흡수모듈로부터 기술정보를 입수하고 기술정보와 시스템 정보가 수록된 데이터베이스를 관리·갱신한다. 또한 접근모듈로부터 질문을 받아 그 결과 세트를 생성하는 기능도 수행한다. 접근모듈은 소비자의 정보 요청 및 입수를 지원하는 서비스와 기능을 수행한다. 즉 소비자의 요청을 접수하고 이에 따라 데이터 관리모듈로부터 기술정보를, 아카이브 보존모듈로부터 보존정보패키지를 입수하여 이용자의 질문과 요구사항에 적합한 배포정보패키지를 생성하고 사용자에게 전송하는 작업을 수행한다. 또한 특별히 보호해야 하는 정보에 대한 접근제한 기능을 수행하며 요청작업과 전달작업이 성공적으로 이루어지도록 하는 조정기능도 수행된다.

운영모듈은 아카이브 시스템의 전반적인 운영을 위한 서비스와 기능을 수행한다. 이를 위하여 정보자원 제출관련 생산자와의 정책 협상, OAIS의 핵심 모듈에 대한 정기적 점검 및 업데이트, 아카이브 관련 정책과 실행절차 점검 및 업데이트, 이용자 요구사항 점검 및

모니터링과 같은 기능을 실행한다. 특히 시스템의 하드웨어와 소프트웨어 사양관리와 아카이브에 구축된 콘텐츠의 무결성을 점검하고 정기적으로 업데이트하는 기능도 수행한다. 보존계획모듈은 OAIS 환경을 감독하고 이용자 커뮤니티의 장기적 접근이 가능하도록 하는 권고안 또는 계획을 제안하는 기능과 최근의 기술변화를 점검하고 새롭게 출현하는 디지털 기술 정보표준, 컴퓨팅 플랫폼 등을 점검하는 기능을 담당한다. 또 생산자, 이용자 커뮤니티, 정보기술에 대한 최근 경향과 변화를 조사 파악하여 현 보존시스템을 수정 업그레이드하고 향후 중장기적으로 적절한 보존전략과 표준을 제안하며 새로운 정보패키지 디자인과 상세한 마이그레이션 계획과 프로토타입을 개발하는 일을 담당한다(그림 1 참조).

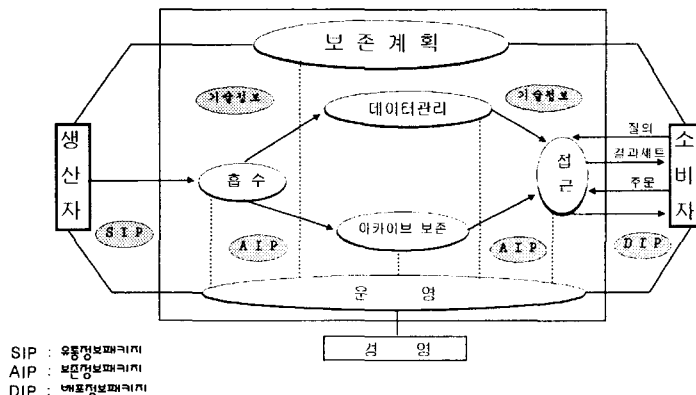
OAIS 기능모형은 타 기관이 참조할 수 있는 기본적인 틀을 제안하는 것이므로 보존정책 수행에 있어서는 해당기관과 조직의 정책에 따라 다양하게 적용할 수 있고 SIP, AIP, DIP의 논리적 형태와 각 모듈의 세부기능 또한 디지털 아카이브의 환경에 따라 다를 수도

있고 다를 수도 있다(Holdsworth 2001).

2. 2 InterPARES의 보존기능 모형

InterPARES는 International Research on Permanent Authentic Records in Electronic Systems의 약어로서 전자문서의 영구 보존을 위한 정책개발 및 모형구축을 하기 위한 대규모 정부·산학공동연구프로젝트를 가리킨다. 특히 실증적이고 귀납적인 방법과 이론적이고 연역적인 방법을 이용하여 전자문서의 진본성 유지, 전자문서의 평가, 진본 전자문서의 장기보존에 대한 정책과 모형을 심층적으로 제시하였다.

InterPARES에서 제시한 보존기능 모형은 OAIS 참조모형에 준하여 제안되었으나 그 범위와 기능은 훨씬 한정적이다. 이는 디지털 정보자원을 전자기록물로 한정시켰고 보존정보에 대한 이용자와 시스템간 교류와 관련된 기능을 포함시키지 않았기 때문이다. 그 반면 전자기록 보존활동에 필요한 업무의 정의와 흐름을 상세히 제안하였다(Thilbodeau 2001).



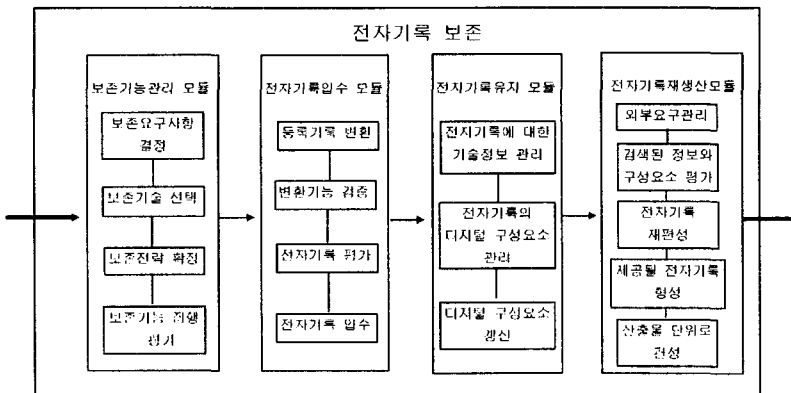
<그림 1> OAIS 기능 모형

〈그림 2〉는 InterPARES 연구가 제안한 전자기록 보존기능을 간단하게 도식화시킨 결과이다. 보존기능은 크게 4가지 모듈로 구성되어 있다: 그 첫 번째 모듈은 보존전략, 보존계획, 보존방법, 이를 위한 기술적 하부구조 등을 포괄적으로 결정하는 '보존기능관리' (Managing the Preservation Function) 모듈이고, 두 번째 모듈은 보존을 위하여 선정된 기록물을 흡수하여 보존방식에 적용시키는 '전자기록입수' (Bringing Electronic Records) 모듈이다. 세 번째 모듈은 선정된 기록물에 대하여 이미 제시된 보존전략 방식과 보존방법을 적용시켜 유지·관리하는 '전자기록유지' (Maintaining Electronic Records) 모듈이다. 마지막 모듈은 '전자기록재생산' (Reproducing Electronic Records) 모듈로 외부의 요구에 응하여 보존된 진본 전자기록을 제공하기 위한 기능들로 이루어져 있다.

2. 3 IRMS의 보존기능 아키텍처

디지털 아카이빙은 기존의 정보자원관리시

스템(IRMS)에서도 중요한 이슈로 부각되어 현용시스템과 아카이브, 현 문헌과 보존문헌 간의 연계를 처리할 수 있는 시스템이 개발되고 있다. 즉 이용자가 아카이브나 정보저장소(repository)에 따로 보존되고 있는 디지털 정보자원을 현용시스템에서 신속하게 접근하여 검색할 수 있을 뿐만 아니라 원하는 포맷과 형식으로 정보자원을 받아볼 수 있는 통합 시스템이 구축되고 있다. Chen(2001)은 아카이브와 연계된 현행의 정보자원관리시스템에서 다음과 같은 세 가지 기능이 구현될 때 정보자원이 효과적으로 관리되고 장기적으로 보존될 수 있다고 하였다: 첫째, 다양한 보존기술(예 매체변환, 마이그레이션, 에뮬레이션 등)들을 필요시 자동적으로 적용시켜 디지털 정보자원을 장기적으로 유지시키는 기능, 둘째, 다양한 포맷과 방법으로 보존되고 있는 데이터를 검색·유통시킬 수 있는 기술적 인터페이스 기능, 마지막으로 이용자의 활용을 수월하게 해주는 이용자편의 인터페이스 기능이다. 다음 〈그림 3〉은 이와 같은 세가지 기능이 정보자원관리시스템과 연계되어 이루어지는 아



〈그림 2〉 InterPARES 보존기능 모형

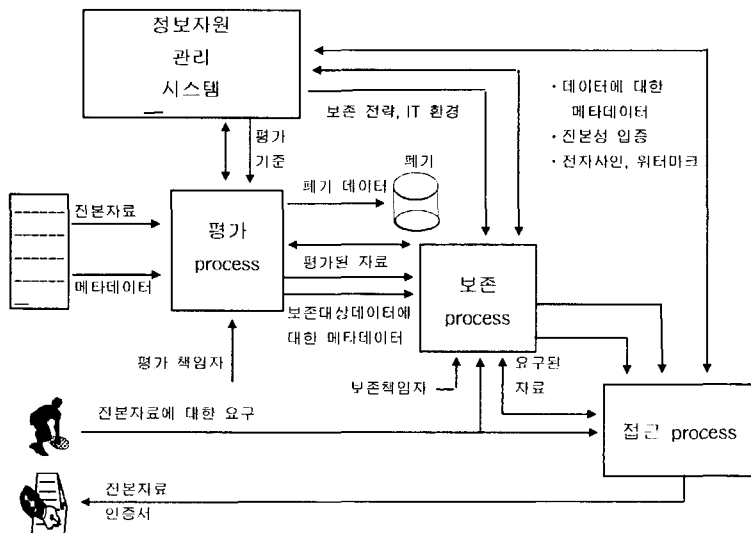
키텍처를 보여주고 있다. 통합시스템은 평가 프로세스에서 현용자원 중 보존할 가치가 있는 정보자원만을 선정하여 메타데이터와 함께 아카이브에 보존시키고, 접근 인터페이스를 통해 이용자가 보존정보자원에 접근 및 검색이 가능하도록 해준다. 또한 정보자원이 계속적으로 진본형태로 보존될 수 있도록 하는 보안과 처리기능도 갖추고 있다.

2. 4 NAL 디지털 아카이브 아키텍처

미국의 국립농업도서관(NAL: National Agricultural Library)은 1997년에 USDA (United States Department of Agriculture) 디지털 출판물에 대한 보존계획을 수립하여 일반 이용자들에게 디지털 정보자원을 이용할 수 있게 관리해오고 있었으나(Uhlir 1997), 최근 디지털 정보자원(자체적으로 인쇄형 정보자원을 디지털포맷으로 변환시키거

나 원래 디지털 형태로 생성된 정보자원)이 급속히 증가되자 보다 체계적이며 영구적인 디지털 아카이빙을 수행하기 위하여 디지털 아카이브를 구축하기로 결정하였다(Hodge 2002). 현재 개발한 프로토타입의 NAL 디지털 아카이브는 OAI 참조모형이 권고하는 프로토콜, 표준, 정책을 따르고 있으며 아카이브의 세부기능은 <그림 4>와 같다.

NAL 디지털 아카이브는 먼저 자료들을 디지털 정보저장소에 축적시키는 작업을 수행한다. 이를 위하여 자생적 전자형태의 USDA 자료들은 변환작업 없이 NAL 저장소에 보내져 등록시키는 반면, 인쇄형 USDA 정보자원인 경우에는 문자인식기를 이용하거나 이미지 스캔 방식을 통하여 디지털화를 시킨 다음, '전자 제본' 기능을 통하여 축적, 검색 가능하도록 변환시킨 후 NAL 저장소로 보내진다. 즉 인쇄자료는 XML문헌, 이미지, 또는 아스키 파일로 변환되어 저장소에 등록



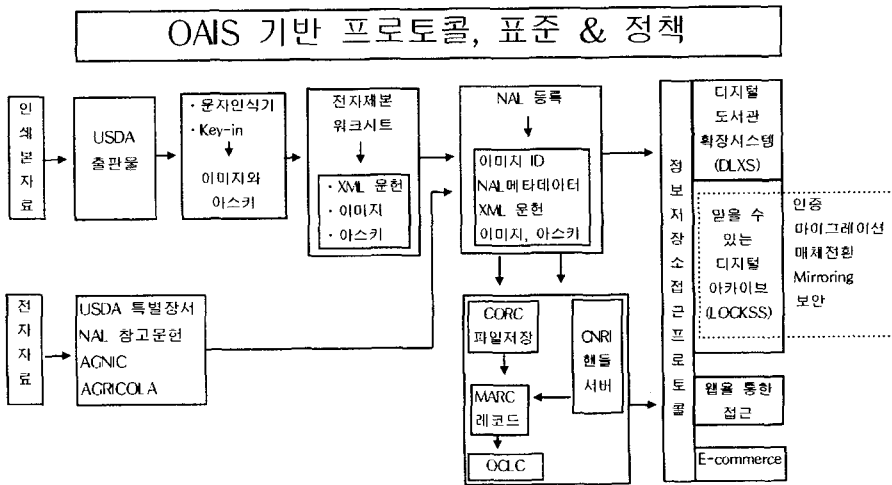
<그림 3> IRMS의 보존기능 아키텍처

축적되어진다. 다음 단계로 저장소에 입수된 각각의 정보자원에 대해 NAL 형식의 메타데이터를 첨가시키고 또한 CORC¹⁾을 이용하여 각각의 자원에 대한 MARC 데이터를 생성한 후 이 데이터를 OCLC에게 보낸다. 이때 각 데이터는 CNRI(Corporation for National Research Initiative)가 개발한 핸들서버에 의하여 고유 식별자가 관리되어지도록 하고 있다. 이와 같이 축적된 정보자원을 확장된 디지털 도서관시스템(DLXS)이나 웹기반 시스템이 용이하게 접근할 수 있도록 저장소 접근프로토콜 인터페이스(repository access protocol)를 갖추도록 하고 있다. 아직까지는 NAL 디지털 아카이브는 USDA 정보자원을

축적, 보존하는데 힘쓰고 있으나 미래에는 보다 강력한 디지털 도서관 시스템과 LOCKSS²⁾과 같은 기술을 바탕으로 효율적이고 활용성이 높은 디지털 아카이브를 구현할 것으로 보인다. 또한 신뢰성 높은 디지털 아카이브를 구축하기 위하여 인증작업, 보존기술, 보안기능 등을 시스템에 활용시킬 계획도 가지고 있다(Frangakis 2002).

2. 5 NEDLIB의 디지털 아카이브 아키텍처

8개의 유럽국가 도서관³⁾은 통합 디지털 저장소 시스템을 구축하기 위하여 NEDLIB(Networked European Deposit Library) 프로젝트



<그림 4> NAL 디지털 아카이브 아키텍처

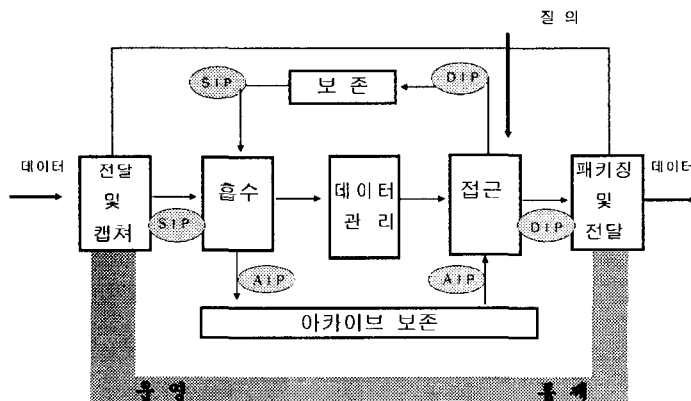
- 1) Cooperative Online Resource Catalog의 약어인 CORC는 OCLC가 지원하는 사업으로 디지털 정보자원에 대하여 표준포맷으로 기술한 후 축적하도록 하여 이용자가 쉽게 검색할 수 있게 해주는 환경을 제공하는 시스템이다.
- 2) LOCKSS는 Lots of Copies Keep Stuff Safe의 약어로 웹 출판 과학학술잡지에 다양한 이용자가 접근이 가능하게 하는 기능으로 도서관을 위하여 디자인된 도구이다. 즉 LOCKSS는 도서관이 용이하게 구독하는 잡지들을 관리하고 (구독하는 잡지가 인쇄형이든 전자포맷이든 상관없이) 검색하고, 계속해서 보존할 수 있도록 해주는 프로그램이다(Reich & Rosenthal 2001).
- 3) 프랑스, 노르웨이, 핀란드, 독일, 포르투갈, 스위스, 이탈리아, 네덜란드

트를 수행하였다. 이 프로젝트의 목적은 도서관에서 활용될 디지털 정보자원의 장기적인 보존에 있으며, 구축되는 디지털 정보자원 보존시스템을 단일시스템(Stand-alone) 형태로 구축하는 것이 아니라 기존에 개발된 디지털 도서관 기반구조에 끼워 넣어 질 수 있는 플러그 인(Plug-in) 모델로 구축되었다는 것이 특징이다. 따라서 제일 먼저 수행된 작업은 디지털 도서관 하부구조 아래 독립된 모듈로써 OAIS표준에 부합하는 디지털 저장소를 설계·구축하는 일이었다. 현재 DAIS(Digital Archival Information System) 버전 1.0이 IBM에 의하여 구축되었다(Werf 2002).

〈그림 5〉는 DAIS의 아키텍처로 OAIS 참조모델에 제시된 흡수, 아카이브 보존, 접근, 데이터관리, 운영 기능이 포함되어있고 그 외 전달과 캡처(Delivery & Capture)와 패키징과 전달(Packaging & Delivery)기능 모듈이 추가되었다. 이 두 모듈은 기존 도서관 시스템과 인터페이스 하는 기능으로 첫 번째 모듈은 여러 가지 파일포맷으로 형성된 기존 데이터를 아카이브에 불러오는 작업이며, 두 번째 모

듈은 이용자에게 적합한 형태로 변환시켜 보내주는 작업을 뜻한다.

아키텍처의 기능에 따른 정보의 흐름을 살펴보면 데이터가 전달과 캡처 모듈에 의하여 유통정보패키지가 생성되면 흡수(Ingest) 모듈에 보내지게 되고 여기서 보존정보패키지가 생성되어 함께 그 데이터가 아카이브 보존(Archival Storage)모듈에 저장되어진다. 또한 아카이브 보존 모듈에는 백업과 재생 기능, 재해복구기능 등이 수행되어지고 데이터 관리(Data Management) 모듈에서 기술 메타데이터(descriptive metadata)가 관리되어진다. 한편 접근(Access) 모듈에 질의가 입력되면 아카이브 보존 모듈로부터 적합한 보존정보패키지를 불러들여 배포정보패키지로 변환시킨 다음 패키징과 전달 모듈로 보내진다. 이로써 이용자에게 적합한 형태로 데이터를 제공할 수 있도록 한다. 또한 DAIS는 배포정보패키지를 저장하여 또 다른 형태의 데이터로 활용시키는 보존(preservation) 모듈, DAIS 시스템 전체를 관리해주는 운영(Administration) 모듈, 통제(Monitoring & Logging) 모듈을 두



〈그림 5〉 DAIS 아키텍처

고 있다.

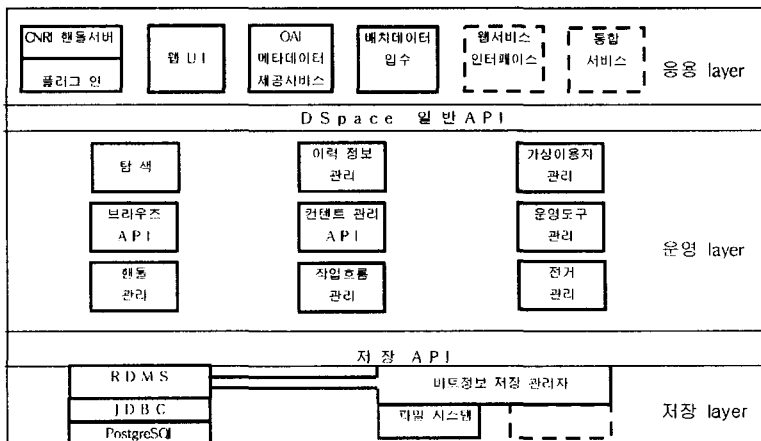
2. 6 DSpace 아키텍처

2000년 3월에 Hewlett-Packard사와 MIT도서관은 다양한 디지털 포맷으로 형성된 지적 산출물에 대한 역동적인 정보저장소 DSpace를 개발하기 시작하여 2002년 11월에 구현시켰다. 특히 DSpace 연구팀은 다른 디지털 아카이브보다 디지털보존과 접근기능을 향상시키는데 주력하였고 동시에 디지털 저작권 관리기능 지원이 가능하도록 개발하였고 DSpace 시스템을 개방형 구조로 설계하여 운영체계가 다른 기관에서도 바로 적용되며 자관요구에 맞추어 수정·확장이 가능하도록 개발하였다(Smith et al. 2003).

〈그림 6〉에 나타난 DSpace 아키텍처를 살펴보면 먼저 크게 세 개의 층 즉 '저장', '운영', '응용' 층으로 구성되어 있는 것을 알 수 있다. 또한 각 층 사이에는 앞으로 있을 기능변경과 성능강화를 위하여 응용 프로그래밍 인터페이스(API: Application Programming Interface) 기능을 두고 있다.

'저장' 층은 데이터 축적 및 데이터관리에 필요한 소프트웨어들로 구성되어있고, '운영' 층은 DSpace에서 이루어지는 모든 기능을 수행할 수 있도록 콘텐츠관리 모듈, 작업흐름관리 모듈, 탐색 모듈, 브라우저 모듈, 운영도구 모듈 이용자관리 모듈 등으로 구성되어 있다. '응용' 층에 있는 각각의 모듈들은 DSpace를 다른 시스템에 적용시킬 때 기능을 대체시키거나 강화시킬 수 있도록 하는 API 기능을 가지도록 하였다. 마지막으로 '응용' 층은 웹UI와 배치형로더(batch loader), OAI(Open Archive Initiative) 프로토콜 지원기능과 핸들서버 기능과 같은 인터페이스를 제공한다. 앞으로 상호운영을 지원하는 웹서비스 인터페이스와 DSpace를 운영하는 여러 기관과의 연계를 지원하는 통합서비스가 추가될 것으로 보인다.

DSpace에서는 디지털 데이터를 영구적으로 보존하는 방법으로 '비트보존'(bit preservation)과 '기능적 보존'(functional preservation)을 제공한다.



〈그림 6〉 DSpace 아키텍처

vation) 방법 두 가지를 채택하여 이용하고 있다. 즉 원래의 포맷을 계속 유지·보존 시켜주는 비트보존 방식과 포맷기술, 매체, 기술 패러다임이 진화됨에 따라 기술변환을 적용시켜 그 내용을 유지·보존시켜주는 기능적 보존방식 두 가지를 다 사용한다. 현재 기능적 보존을 위하여 DSpace에서 자원해주는 포맷은 TIFF, AIFF, XML, PDF, RIFF이다.

2. 7 Knowledge Bank 아키텍처

미국의 오하이오 주립대학에서 이루어지고 있는 가상교육과 평생교육을 위하여 축적해 놓은 지적컨텐츠를 보다 효율적으로 관리하고 활용할 수 있게 해주는 지식저장소의 필요성이 대두되자, 이에 2002년 OSU대학의 디지털 정보자원 저장소를 구축하기 위한 Knowledge Bank 프로젝트가 시작되었다(Falk 2003). Knowledge Banks는 대학의 중요하고 가치 있는 지식컨텐츠를 선정하여 단순히 축적하는 것이 아니라 이러한 지식컨텐츠를 지적자산으로 형성되어 활용될 수 있도록 하는 디지털 아카이브이라고 말할 수 있다. 특히 Knowledge Bank는 다양한 형태의 디지털 정보자원에 대한 검색, 소장 및 보존기능과 다양하고 강력한 접근기능을 지닌 사용자 인터페이스를 제공하며 나아가서 지식관리까지 수행할 수 있는 지능적 디지털 아카이브라 할 수 있다. Knowledge Bank가 제공하는 기본적 기능을 살펴보면 다음과 같다(Branin 2002).

- 지식객체를 유통시키고 기술해주는 인터페이스

- 디지털 지식객체와 메타데이터를 유지시켜 주는 하부 시스템
- 기능적 하부구조:
 - 색인기능
 - 식별기능
 - 콘텐츠 및 시스템 관리기능
 - 탐색 및 검색 기능
 - 보존기능
 - 프로토콜과 포맷변환기능
 - 이용자확인과 인증기능을 포함한 접근기능
- 지능적 접근, 검색, 통합기능을 지원하는 이용자 인터페이스

Knowledge Bank도 DSpace와 같이 크게 세 개의 층으로 이루어지고 있다. 첫 번째 층은 OSU의 디지털저장소, 웹 정보저장소, 각각 주제별 디지털 정보저장소, 다른 캠퍼스의 디지털저장소 등과 같은 기존의 디지털저장소 그룹을 가리킨다. 이와 같은 다양한 디지털저장소에 소장된 정보자원은 OAI 인터페이스를 기반으로 하여 두 번째 층인 Knowledge Bank 엔진모듈에 보내지게 된다. Knowledge Bank 엔진모듈은 메타데이터 추출기능, 메타데이터와 콘텐츠 디스플레이기능, 보존기능, 확인·인증·청구기능 등을 포함하는 접근관리기능, 콘텐츠와 버전관리기능, 프로토콜 및 포맷변환 기능 등을 수행한다. 특히 이 엔진모듈은 세 번째 층인 OSU Knowledge Bank 이용자 인터페이스와 상호 연결되어 있어 디지털 저장소 즉 Knowledge Bank에 이용자가 원하는 모든 기능을 제공하는 역할을 한다. 이용자 인터페이스 층은 OSU의 지적자

산을 지능적으로 접근할 수 있게 하는 포괄환경을 제공할 뿐만 아니라 이용자가 새로운 데이터를 수월하게 시스템에 보내고 기술할 수 있게 하는 도구도 제공해준다(그림 7 참조).

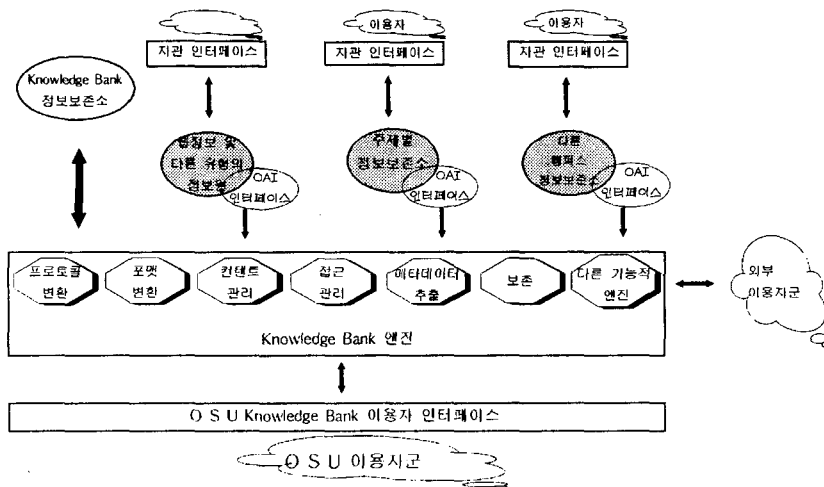
3. 영구적 디지털 아카이브 구축을 위한 전략

디지털 아카이브가 영구적으로 유지되면서 이용자에게 무결성을 지닌 데이터를 제공해 주기 위해서는 아카이브 운영과 체제에 대한 “신뢰”, 디지털 정보자원과 그 활용을 영구적으로 유지해주는 “기술환경 관리”, 시스템과 데이터의 무결성을 확신시켜주는 “보안”이 이루어져야 할 것이다. 본 장에서는 디지털 아카이브가 영구적으로 유지되기 위하여 필요한 세가지 속성에 대하여 살펴보았다.

3. 1 신뢰성 높은 디지털 아카이브

일반적으로 한 실체는 다른 실체가 자신의 기대한 바와 같이 행동 및 의사결정을 하리라고 가정할 수 있을 때 그 실체가 다른 실체를 신뢰한다고 말할 수 있다(황재훈 외 2002). 따라서 한 디지털 아카이브를 신뢰한다는 것은 정보 이용자 또는 소비자가 가까운 미래에는 물론이고 먼 미래에 이 디지털 아카이브가 계속 유지될 것이며, 아카이브가 제공하는 데이터가 무결성을 지닌 진본일 것이라고 기대하고 또 믿는다는 것을 뜻한다. 이러한 시스템에 대한 신뢰는 “활용” 활동에 상당한 영향을 주는 개념이므로 디지털 아카이브를 운영하고 자 하는 기관은 신뢰성 높은 디지털 아카이브를 구축해야 할 것이다.

RLG/OCLC 공동연구그룹(2002)은 신뢰성 높은 디지털 아카이브가 지녀야 되는 속성 7가지를 제안하였다. 따라서 디지털 아카이브의 신뢰성을 높이기 위해서는 이와 같은 속성



<그림 7> Knowledge Bank 아키텍처

을 갖출 수 있도록 개발되어야 할 것이다.

1) OAIS 참조모형에 기반을 둔 시스템

아카이브의 신뢰성을 높이기 위해서는 먼저 디지털 아카이브의 표준화를 위하여 제안된 OAIS 참조모형이 제시하는 환경과 기능을 따라야 할 것이다. 이로써 디지털 아카이브의 기본적인 구조, 정책, 프로토콜 등이 갖추게 되고 최소한의 기본적 기능을 수행할 수 있는 능력을 지닌다고 판단되어 질 수 있다.

2) 운영책임

디지털 아카이브는 이용자들에게 아카이브 내의 모든 기능이 원활하게 처리되고 있다는 매우 근본적인 신뢰성을 보여주어야 한다. 이를 위해서는 아카이브의 물리적 환경, 데이터 관리, 보안 시스템이 국가/국제적 표준에 적합함을 보여주어야 하고, 수행능력 또한 검증되었음을 알려주어야 한다. 그 외에 운영관리에 대한 신뢰도 인식을 높이기 위해서 시스템 위험관리 및 긴급대책에 대한 성문화된 정책이 수립되어야 한다.

3) 기관의 지속성

디지털 아카이브의 신뢰성을 높이기 위해서는 아카이브가 영구적으로 지속되는 기관임을 입증시켜주어야 한다. 이를 위해 먼저 정책과 편람 등에 문화적, 학술적 자산인 디지털 정보 자원을 장기적으로 보유, 관리하여 계속해서 이용될 수 있도록 하겠다는 아카이브의 임무와 범위를 밝혀야 한다. 또한 아카이브의 정책과 절차는 변화에 따라 개정될 수 있음을 밝혀야 할 것이다.

4) 재정적 지속성

장기적 운영계획 안에서 최선의 실무가 이루어지기 위해서는 재정 면이 안정되어야 하므로 디지털 아카이브는 계속적으로 그리고 충분한 재정적 지원을 받을 수 있도록 노력해야 한다. 따라서 디지털 아카이브 운영에 필요한 단기적 그리고 장기적 재정계획을 수립해야 한다. 특히 장기적 재정계획은 디지털 아카이브의 신뢰를 높여 줄 뿐만 아니라 활동범위와 규모를 추정할 수 있는 기능을 한다.

5) 적용기술과 절차의 적합성

디지털 아카이브는 다양한 보존 전략에 대한 적합성 및 효율성을 공개적으로 논의하고 또 각각의 가능성도 평가하는 기능을 가져야 한다. 즉 아카이브의 모든 관리기능을 수행하는데 있어서 현용의 하드웨어와 소프트웨어가 적합한지를 자체적으로 또는 외부에 의하여 평가해야 하며 필요시 기술전환을 할 수 있게 하는 정책과 계획을 가지고 있어야 한다. 따라서 아카이브의 직원들은 구축된 시스템의 기술과 절차에 대하여 정확히 이해하고 그 성능을 평가할 수 있어야 한다.

6) 시스템 보안

신뢰성 높은 디지털 아카이브는 디지털 자산이 보호받도록 설계되어야 한다. 따라서 아카이브에서 이루어지는 모든 디지털 자원의 복사과정, 인증과정, 방화벽시스템 및 백업시스템 등에 관한 정책과 절차가 이용자 요구사항과 부합되게 개발되어야 하고, 재난시 그에 따른 처리 및 복구방법에 대한 정책과 계획도 세워야 한다. 특히 데이터의 무결성을 유지하

기 위하여 데이터 손실, 데이터 변경, 변경된 데이터의 재저장 등을 방지할 수 있는 방법과 이러한 변화에 대한 감지 방법에 대한 계획을 갖추어야 한다.

7) 절차적 책임

디지털 아카이브는 그 목적에 맞게 수립된 정책과 절차에 따라 업무를 수행해야 하는 책임을 가지고 있다. 따라서 아카이브의 모든 업무절차는 성문화가 되어야 하고 통제 메커니즘을 통하여 평가되어 피드백이 될 수 있도록 해야 한다.

이와 같은 7가지 속성은 실제 디지털 아카이브 기능면보다는 운영의 신뢰도를 높여주는 것으로 디지털 아카이브가 구축되어 운영되기 전 미리 갖추어져야 할 것이다. 또한 디지털 아카이브의 신뢰도를 높이기 위해서 RLG/OCLC 연구진이 제시한 조건 이외에의 또 다른 방안을 다음과 같이 들 수 있다: 첫째, 인증 받은 교육기관에서 육성된(즉 교육받고 훈련받은) 기록관리전문가, 정보전문가, 그리고 관리자들이 혼합하여 디지털 아카이브를 운영하고 각각의 업무를 수행할 때 높은 신뢰도를 얻을 수 있다. 둘째, ISO 9000, DoD5015.2-SID, 공공기록청 표준(영국), BSI DISC0008 등과 같은 기준에 의하여 디지털 아카이브의 절차와 과정에 대한 평가를 받아 업무수행 인증을 받도록 하는 방법이다. 이러한 ISO 인증은 기관의 신뢰를 높여줄 수 있다. 셋째, 디지털 아카이브에서 제공되어지는 데이터가 진본이며 무결성을 지닌다는 것을 증명해주는 “인증” 작업을 수행하는 방법이다. 특히 개개인에 대한 인증서 발급은 시스템과 데이터의 안전과 보

안을 증명해주는 기능도 하므로 아카이브의 신뢰도를 높일 수 있다.

3. 2 지속성 높은 디지털 아카이브

디지털 아카이브가 정보자원을 영구적으로 보존하기 위해서는 정보기술 발전에 발맞추어 새로운 기술의 선택, 적용, 평가 업무와 같은 기술관리 능력이 필요하다. 따라서 시스템 스스로 그리고 효율적으로 기술관리(technology management)를 수행할 수 있는 영구적 아카이브(persistent archive)가 구축되어야 할 것이다. 영구적 아카이브는 아카이브 기능의 점진적 변화와 콘텐츠의 장기적 유지를 위하여 지능적 아카이빙 스킴을 도입하여 외부적, 내부적 변화를 아카이브 시스템 자체적으로 인지하고, 필요시 새로운 기술이나 방법을 자동적으로 적용시켜 시스템이 스스로 진화되도록 함으로써 디지털 아카이브의 생명력을 높여주는 방식을 말한다(Marcum 2002). 특히 저장된 정보자원이 미래에도 계속해서 검색되고 검색된 데이터는 미래의 새로운 소프트웨어에서도 작동될 수 있도록 하는 기능을 가져야 영구적 아카이브라 할 수 있다.

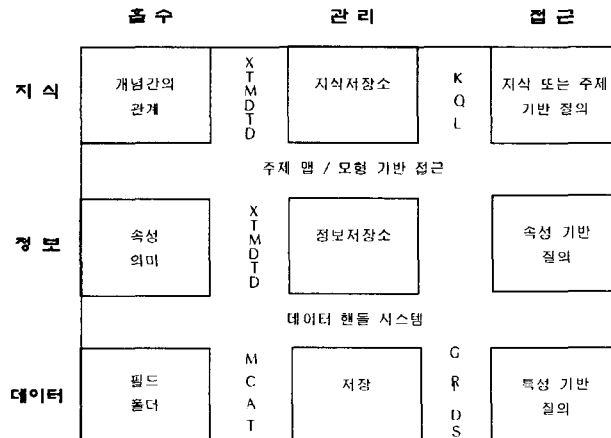
영구적 아카이브의 개발은 DARFA, NSF, NARA의 지원을 받아 San Diego 슈퍼컴퓨터 센터에 의하여 이루어졌다(Moore 2001a). 이들은 디지털 아카이브가 ‘영구적’이 되기 위해서는 디지털 아카이브를 구성하고 있는 정보자원 객체와 기술적 아키텍처, 이 두 가지가 환경과 요구에 맞게 자동적으로 진화되어야 한다는 전제 아래, 정보자원 객체를 정보기술 진화에도 영향을 받지 않는 영구적 객체 포맷

으로 변환시켜 보존하는 방법과 시스템의 하드웨어와 소프트웨어를 점진적으로 진화시키면서 그 충격을 최소화시킬 수 있는 방법을 개념적으로 제안하였다. 즉 이들은 정보자원의 퇴화 문제를 해결하는 방법뿐만 아니라 장기보존 아카이빙 기능을 수행할 수 있는 아카이브 아키텍처를 제안하였다.

〈그림 8〉에서 보듯이 San Diego 슈퍼컴퓨터 센터 연구팀은 정보자원 객체를 기술하는 단계를 세 가지로 나누었고(데이터, 정보, 지식) 이러한 정보자원을 관리하는 기능을 역시 세 가지로 나눈 후(흡수, 관리, 접근), 각각의 세단계의 모드가 만들어내는 9가지 그리드에서 영구적 아카이브가 수행해야할 관리기술과 기술정보 범주를 제안하였다. 즉 제안된 영구적 아카이브는 먼저 흡수된 데이터 객체에 대하여 물리적, 논리적, 개념적 속성을 분석한 후, 이를 토대로 하여 기술정보를 형성한다. 다음 비트스트링 형태로 존재하고 있는 데이터 객체에 태그형식으로 그 기술정보를 첨가시킨다. 이러한 태그는 비트스트링으로 존재

하는 정보자원의 속성을 알려주는 중요한 정보이며 모든 데이터가 유기적으로 연관될 수 있도록 해준다. 또한 태그정보는 데이터 모델, 데이터 엘리먼트 정의, DTD, 스타일 정의, 온톨로지, 텍사노미, 시소러스, 주제 맵, 추론규칙 등과 같은 정의에 의하여 비트스트링이 다시 표현될 수 있게 해주는 중개자 역할을 해준다.

또 한편, 영구적 아카이브의 아키텍처는 OAIS표준에서 권고한 세 가지 기본 기능 즉 입수, 관리, 접근을 수행할 수 있도록 구성되어 있다. 입수 과정에서 퇴화된 포맷으로 구성된 데이터객체는 파싱과 태깅 작업을 걸쳐 영구적 포맷으로 변환되어 저장되어지고, 저장된 데이터는 지식기반 운영과정을 통하여 관리되어진다. 보존된 정보자원에 대한 접근요구가 있을 경우 적합데이터는 검색되고 검색된 데이터객체는 현 기술에 적합한 형태로 변환되어진다. 이러한 데이터 변환(materialization)은 영구적 포맷에서 현재 사용하고 있는 기술에 적합한 형태로 변환시켜주는 것을 의미한



〈그림 8〉 영구적 아카이브의 아키텍처와 인터페이스

다. 이러한 과정을 걸쳐 디지털 정보자원이 미래에도 보존되고 활용될 수 있음을 보여주고 있다.

각 단계별로 이루어지는 9가지의 과정이 보다 수월하게 진행될 수 있도록 시스템은 데이터의 필드와 그 관계를 판단해주는 함축 지식(implied knowledge), 데이터의 구조정보를 알려주는 구조지식(structural knowledge), 주제적 개념간의 관계를 파악하게 해주는 도메인지식(domain knowledge), 디지털객체의 생성과 관련된 절차지식(procedural knowledge), 지식기반 질의를 지원해주는 표현지식(presentation knowledge)을 구축하여 지식기반 아카이빙을 할 수 있도록 하였고 특히 9개의 과정들이 잘 연결될 수 있도록 모든 과정 사이에는 소프트웨어 중개자인 인터페이스를 두도록 하여 각각의 과정들이 별 문제없이 연결되도록 하였다(Moore 2002b).

San Diego 슈퍼컴퓨터 센터의 연구팀이 개발한 영구적 아카이브가 표준이며 완성품이라 할 수는 없지만, 지속성 높은 디지털 아카이브를 구축하기 위해서는 이와 같은 기능을 수행할 수 있는 영구적 아카이브가 필요하다는 것은 분명하다.

3.3 보안성 높은 디지털 아카이브

디지털 정보자원이 위조, 변경, 훼손, 손실되지 않고 계속 보존되기 위해서는 보안성 높은 디지털 아카이브 즉, 보안 메커니즘을 지닌 디지털 아카이브가 개발되어야 한다. 아카이빙 데이터를 보호하기 위해서는 일반적으로 두 가지 메커니즘을 이용하고 있다. 첫 번째

방식은 전통적으로 사용된 것으로 사용자 ID와 비밀번호를 이용하여 접근을 통제하는 방식을 말한다. 또 다른 방식은 전자서명 기술을 활용하는 것으로 현재 공개키 암호기술에 기반을 두고 정보요구 당사자나 정보전달 상대방의 신원을 확인하여 접근을 통제하는 방식을 말한다.

공개키 암호기술에 기반을 두는 방식은 사용자가 자신만이 알고 있는 전자서명생성키를 이용하여 수학적인 연산을 통하여 자신만의 고유한 전자서명 값을 계산한 후, 그 결과를 수신자에게 송신한다. 수신자는 송신자가 제공하는 전자서명검증키를 사용하여 전자서명 값의 진위 여부를 다시 수학적인 연산으로 확인한 후, 올바른 결과 값이 나오는 경우에만 디지털 정보를 접수하는 방식을 말한다(염홍열 & 홍기용 1998). 이때 본인의 공개키를 악용하는 것을 방지하기 위하여 그 공개키가 본인의 것임을 확인하는 인증기능이 필요한데 이를 해결하기 위하여 즉, 공개키의 무결성을 보장하기 위해 “공개키기반구조”(PKI: Public-Key Infrastructure) 방식을 활용하고 있다. PKI는 공개키를 공개하는 대신 공개키와 그 공개키의 소유자를 연결하여 주는 인증서(certificate)만을 공개하여 신뢰객체가 아닌 사람은 그 문서의 내용을 변경할 수 없도록 함으로서 디지털 데이터의 무결성을 보장할 수 있게 해주는 방법이다(황재훈외 2000).

디지털 아카이브의 보안성을 높여주기 위해서는 이제 단순한 ID와 비밀번호를 통한 접근 허용 방식이 아닌 인증서를 받은 신뢰객체만이 디지털 정보자원에 접근 활용할 수 있도록 하는 PKI 기반 유통구조를 구축하여야 할 것

이다. 이로써 정보의 기밀성과 무결성이 유지되고 보다 안정된 전송이 보장될 수 있다. 그러나 PKI 기반 유통구조는 이용자가 비밀키를 관리, 보안, 저장, 활용하는 책임을 가지도록 하는 점이 문제점으로 대두되고 있다. 일반적으로 최종이용자들은 이러한 암호키의 보안이 왜 중요한지에 대한 인식이 부족하고 하나의 암호키를 관리하는 것도 어려워하기 때문이다(Ellison & Schnieier 2000). 이러한 키 관리 문제를 해결하는 방법으로는 암호키를 내포하고 있는 “스마트카드(smart card)”를 활용하는 방안이 있다. 스마트카드를 사용자와 스마트카드의 소유자가 일치한다면 최종이용자의 키관리 문제는 쉽게 해결될 수 있을 것이다.

또한 디지털 아카이브는 현재 사용하고 있는 보안 메커니즘을 계속적으로 관리 유지하기 위해서는 암호화시스템과 인증시스템에 대한 보안 메타데이터를 작성해야 한다. 이것은 하나의 매체를 계속 사용하기 위해서는 그 매체를 사용하고 있는 하드웨어에 대한 속성을 기술하듯이 데이터 보안관련 하드웨어/소프트웨어에 대한 속성도 기술되어야 할 것이다. 보안 메타데이터는 다음과 같은 항목에서 기술되어 질 수 있다(Schroeder & Perrine 2001).

- 1) 인증(Authentication) 속성
 - 원하는 강도
 - 종류
 - 종류 상세설명
 - 데이터
- 2) 관허(Authorization) 속성
 - 원 소유자

- 접근통제 리스트
- 3) 암호화(Encryption) 속성
 - 원하는 강도
 - 암호키
 - 4) 민감도(Sensitivity) 표시
 - 5) 등급하락(Downgrade) 정보
 - 6) 아카이브 활동에 대한 기록

또 한편으로 분산환경에서 서버와 서버, 클라이언트와 서버간의 연결에 있어서의 신뢰를 높여주고 서로 보호해주기 위하여 연결된 이용자와 서버를 인증해주는 그리드 보안구조(GSI: Grid Security Infrastructure)를 디지털 아카이브의 네트워크 환경에 적용시키고 있다. 특히 GSI는 웹환경에서 연결되는 클라이언트/서버간의 관계를 인증하여 웹을 통하여 전달되는 다양한 데이터의 신뢰를 확신시켜주는 기능을 하므로 웹환경에서 구축되는 디지털 아카이브에서는 GSI 보안 메커니즘 활용도 고려해야 할 것이다.

4. 디지털 아카이브 개념적 모형 설계

디지털 아카이브는 디지털 정보자원을 체계적으로 보존시켜 미래에도 활용될 수 있도록 정보자원의 평가 및 입수, 소장 및 보존, 접근과 같은 처리기능과 네트워크 환경, 정보기술 환경, 생산자와 이용자 환경 등까지도 관리하는 전반적인 시스템을 말한다. 현재 디지털 아카이브는 2002년 ISO 표준으로 공표된 OAIS 참조모형에 근거하여 구축됨에 따라 기본적 기

능과 절차가 갖추게 되었고 정보자원을 서술하는 기술정보와 기반환경이 표준화가 됨에 따라 아카이브간의 상호운영 및 자원공용이 가능해졌다. 다만 OAIS 참조모형은 개념적으로 최상위 차원에서 정의 내려진 것이기 때문에 각각의 디지털 아카이브는 제반 환경과 목적에 맞게 수정 보완하여 개발되어야 한다. 따라서 본 장에서는 디지털 정보자원을 장기적으로 또는 영구적으로 보존하기 위해서는 디지털 아카이브가 신뢰성과 보안성이 높아야 하고 지능적 관리기술이 필요하다는 전제아래, OAIS 기능모형이 제안하는 6가지 기본 기능을 기초로 하여 보안과 통제, 지능형 관리, 지식기반 처리기술과 같은 기능이 부가된 디지털 아카이브 기능모형을 제안하였다. 제안된 디지털 아카이브는 다음과 같이 6개의 모듈로 구성되어 있다(그림 9 참조).

1) 흡수모듈

인쇄 및 전자자료 또는 이미 다른 기관에서 만들어진 SIP가 아카이브에 들어오면, 아카이브의 목적과 데이터의 가치를 기초로 하여 보존가치 여부를 평가한 후 '보존'하기로 결정된 인쇄자료인 경우에는 적합한 방식을 이용하여 디지털 자료로 변환시키고 전자자료이거나 SIP인 경우에는 흡수과정에서 품질변화가 없었는지를, 즉 진본성과 품질을 확인한다. 진본이 확인된 정보자원에 대하여 개념적·물리적 분석을 하여 기술정보를 추출하고 이러한 기술정보는 데이터관리 모듈로 보내진다. 한편 진본 디지털 정보자원이 영구적으로 보존될 수 있도록 보존포맷인 AIP 형태로 변환시킨 후 보존모듈로 보낸다.

2) 보존모듈

보존모듈은 흡수모듈로부터 AIP를 입수하여 장기 축적시킨 다음, 접근모듈로부터 요청이 있을 경우 저장된 AIP의 복제본을 전송시키는 일을 담당한다. 이러한 데이터를 장기보존하기 위해서는 정기적으로 새로운 보존기술을 적용시키고, 하드웨어와 소프트웨어를 변환하고, 재난 방지를 위하여 체계적인 위험관리를 실시하는 보존관리 기능을 수행하여야 한다. 즉 보존관리를 통하여 포맷이전, 마이그레이션, 하드웨어 에물레이션, 새로운 저장 매체의 정기적 이전, 지속적인 백업, 물리적으로 분리된 공간에의 복제본 저장, 주기적 오류 점검 등이 이루어 질 수 있다. 이러한 다양한 관리기능을 아카이브 자체적으로 파악하고 수행하기 위해서는 지식기반 관리기능이 갖추어져야 한다.

3) 데이터관리모듈

흡수모듈로부터 기술정보를 입수하여 각 정보자원에 대한 메타데이터 DB와 개념 및 속성데이터 DB를 형성하여 관리하다가, 접근모듈로부터 질문을 받으면 이 두가지 DB를 이용하여 결과를 생성시켜 접근모듈로 보낸다. 메타데이터와 개념·속성데이터는 속성이나 주제를 기반으로 보존정보자원을 검색할 수 있도록 해주고 또한 그 결과를 이용자에게 제공할 때 원하는 포맷형태로 재형성할 수 있도록 해주는 역할을 한다.

4) 접근모듈

이용자에게 요구가 들어오면 먼저 질의처리를 한 다음, 데이터관리모듈로부터 기술정보

를, 보존모듈로부터 AIP를 입수하여 제공포맷인 DIP로 변환하여 이용자에게 전송시키는 역할을 한다. 이때 이용자에게 제공되어지는 데이터가 진본임을 알려주는 인증서가 함께 제공될 수 있도록 인증과정이 이루어져야 한다. 또한 확인된 이용자만이 접근할 수 있게 하는 외부이용자 관리와 이용자의 요구가 잘 처리되었는지 아니면 문제점이 있는지 파악하고 문제점이 있는 경우는 피드백처리를 할 수 있도록 하는 외부요구관리도 이루어져야 한다.

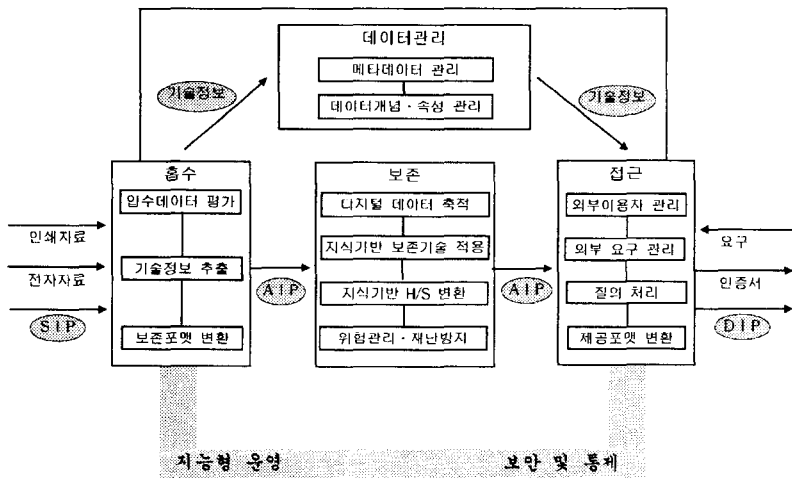
5) 지능형 운영모듈

아카이브가 내외적으로 수행해야 하는 정책 개발과 이에 따라 운영되는 모든 관리과정이 지능형 운영모듈에서 이루어진다. 즉 데이터 입수관련 정책협상, 아카이브 핵심모듈에 대한 정기적 점검 및 업데이트, 아카이브에 대한 이용자와 생산자의 요구사항 점검과 모니터링, 최근 기술변화 점검, 새롭게 출현하는 디지털 표준 및 정보표준 점검, 향후 중장기적 적절한

보존전략과 표준 개발, 새로운 정보패키지 디자인 개발 등의 업무를 맡는다. 이러한 모든 사항에 대한 모니터링 및 관리가 체계적이며 효율적으로 이루어질 수 있도록 지능형 운영 체계를 구축해야 한다.

6) 보안 및 통제모듈

영구적 아카이브 환경을 구축하기 위해서는 시스템 보안과 통제가 필요하다. 이를 위하여 신뢰성 있는 이용자 객체를 확인해주는 스마트카드를 이용하는 PKI 기반 인증 시스템과 아카이브와 네트워크 상에서 연결되는 시스템의 신뢰성을 확인해주는 GSI와 같은 메커니즘을 구축해야 할 것이다. 또한 데이터 자체에 대한 보안이 필요한 경우에는 데이터 암호화를 수행해야한다. 이때, 암호화된 데이터의 영구적 유지를 위하여 데이터 마이그레이션을 정기적으로 수행하듯이 정기적으로 암호 마이그레이션을 수행해야 한다.



<그림 9> 디지털 아카이브 개념적 모형

5. 결 론

디지털 아카이브나 아카이빙 시스템이 효과적으로 구축되기 위해서는 디지털 정보의 보존에 관한 표준화와 그 방법론, 보존기능과 절차를 알려주는 성문화된 정책, 실제적인 안내 지침 등이 필요하다. 이에 OAIS 참조모형이 포괄적이며 상세한 디지털 아카이빙 표준을 제시하였고 이를 반영하여 여러 기관에서 디지털 아카이브를 구축하거나 모형개발 연구를 진행하고 있다. 본 연구는 앞으로 구축될 디지털 아카이브는 OAIS 참조모형이 권고한 기본개념과 기능, 환경을 근간으로 개발된 시스템이어야 하나 보다 신뢰성과 보안성이 높고 정보기술 관리기능이 영구적으로 유지될 수 있는 디지털 아카이브의 개념적 모형을 제시하였다.

지금까지의 디지털 도서관 또는 정보제공기관에서의 정보저장소는 장기보존의 의미는 희박하고 디지털 정보자원의 서비스와 이용에

초점을 맞춘 단기보존의 의미가 강한 저장소라 할 수 있다. 그러나 미래에서도 현재의 소장된 정보자원이 활용될 수 있게 하기 위해서는 디지털 장기보존은 피할 수 없는 디지털 도서관과 정보제공기관들의 임무이며 수행해야 하는 기능인 것이다. 따라서 디지털 도서관 또는 정보제공기관은 현재의 시스템에서 장기보존을 위한 아카이빙 기능을 강화시키거나 아카이브를 구축하여 통합운영을 할 수 있는 방안을 도모 하여야 할 것이다. 또한 디지털 아카이브는 우리의 귀중한 디지털 유산을 장기적으로 보존하고 활용할 수 있게 해주는 기반구조이기 때문에 우리나라 실정에 맞는 국가적으로 작게는 기관별로 디지털 아카이브 구축이나 아카이빙 시스템 개발과 함께 체계적인 디지털 정보자원의 보존계획과 정책이 수립에 더 많은 관심을 보여야 할 것이다. 이를 위해서는 앞으로 더 많은 연구가 여러 분야에서 진행되어야 할 것으로 보인다.

참 고 문 헌

- 김희정. 2003. 전자저널 아카이빙을 위한 OAIS 참조모형의 적용방안에 관한 연구. 『한국기록관리학회지』 3(2): 1-25.
- 서은경. 2003. 디지털 정보자원 보존의 위협관리 분석: 대학도서관 전자정보실 중심으로 『정보관리학회지』 20(1): 5-30.
- 염홍렬, 홍기용. 1998. 공개키 기반구조 『통신정보보호학회지』 8(3): 5-18.
- 이소연. 2002. 디지털 아카이빙의 표준화와 OAIS 참조모형. 『정보관리연구』 33(3): 45-68.
- 정진택. 2001. 디지털 영상 아카이브 구축모형에 관한 연구. 『국회도서관보』 38(1): 20-42.
- 황재훈, 박춘석, 정연식, 송홍엽. 2000. 전자서명을 통한 인증기술과 공개키 기반구조에 대한 고찰. 『산업기술연구소 논문집』 32(2): 105-121.

- Berthon, H., S. Thomas and C. Webb. 2002. "Safekeeping: A Cooperative Approach to Building a Digital Preservation Resources." *D-Lib Magazine*, 8(1).
 <<http://www.dlib.org/dlib/january02/berthon/01berthon.html>>.
- Brainin, J. 2002. *Knowledge Bank Implementation Plan and Support Request*.
 <<http://www.lib.ohio-state.edu/KBinfo/KBImplPlan02.pdf>>.
- Brown, A. 2000. *Digital Archiving Strategy*. English Heritage Center for Archeology.
 <<http://www.ifs.tuwien.ac.at/~aola/links/satekept/e-heritage-strategie.pdf>>.
- CCSDS. 2002. *Reference Model for an Open Archival Information System (OAIS)*.
 <<http://www.ccsds.org/document/pdf/CCSDS-650.0-B1.pdf>>.
- Chen, Su-Shing. 2001. "The Paradox of Digital Preservation." *Computer*.
 <<http://www.gseis.ucls.edu/us-interpares/pdf/ParadoxOfDigitalPreservation.pdf>>.
- Ellison, C. and B. Schneier. 2000. "Ten Risks of PKI: What You're not Being Told about Public Key Infrastructure." *Computer Security Journal*, 16(1): 1-7.
- Falk, Howard. 2003. "Digital Archive Developments." *The Electronic Library*, 21(4): 375-379.
- Frangakis, Evelyn. 2002. *Preserving Digital Publication*.
 <<http://www.nal.usda.gov/preserve/USDADigiPubPrePresProg/ppt/sld013.htm>>
- Hodge, Gail M. 2000. "Best Practices for Digital Archiving: An Information Life Cycle Approach." *D-Lib Magazine*, 6(1).
 <<http://www.dlib.org/dlib/january00/01hodge.html>>.
- Hodge, Gail M. 2002. "Digital Preservation: Overview of Current Developments." *Information Services Use*, 22(2): 73-82.
- Holdsworth, D. 2001. A Blueprint for Representation Information in the OAIS Model.
 <<http://www.personal.leeds.ac.uk/~ecldh/cedars/ieee00.html>>.
- Marcum, Deanna. 2002. "Introduction: The Changing Preservation Landscape." *Conference Proceeding on the State of Digital Preservation: International Perspective*. Washington, D.C: CLIR. pp.1-31.
- Moore, R. W. 2001a. *Knowledge-Based Grids*.
 <<http://www.sdsc.edu/TR/TR-2001-02.doc.pdf>>

- Moore, R. W. 2001b. *Knowledge-Based Persistent Archive*.
<<http://www.sdsc.edu/TR/TR-2001-04.doc.pdf>>
- Reich, V. and D. S. Rosenthal. 2001. "LOCKSS: A Permanent Web Publishing and Access system." *D-Lib Magazine*, 7(6).
<<http://www.dlib.org/dlib/june01/reich/obreich.html>>
- RLG/OCLC Working Group. 2002. *Trusted Digital Repositories: Attributes and Responsibilities*. Mountain View, CA: RLG.
- Rossell, Kelly, 2000. *CEDAR: Longterm Access and Usability of Digital Resources: the Digital Preservation Conundrum*.
<<http://www.ariadne.ac.uk/issue18/cedars>>
- Schroeder, W. and T. Pervine. 2001. Security Models for NARA Electronic Record Management.
<<http://www.sdsc.edu/TR/TR-2001-07.doc.pdf>>
- Smith, Mackenzie et al. 2003. "DSpace: An Open Source Dynamic Digital Repository." *D-Lib Magazine*, 9(1).
<<http://www.dlib.org/dlib/jannary03/smith/01smith.html>>
- Thibodeau, K. et al. 2001. *The Long-term Preservation of Authentic Electronic Records: Finding of the InterPARES Project*.
<<http://www.interpares.org/book/index.cfm>>
- Uhlir, Paul. 1997. *Framework for the Preservation and Permanent Public Access to USDA Digital Publications*.
<<http://preserve.nal.usda.gov:8300/npp/frameprt.html>>
- van der Werf, Titia. 2002. "Experience of the National Library of the Netherlands." *Conference Proceeding on the State of Digital Preservation: International Perspective*. Washington, D.C: CLIR. pp. 54-64.
- Witten, I.H. and D. Bainbridge. 2003. *How to Build a Digital Library*. Amsterdam: Morgan Kaufmann Publishers.