

# 디지털 아카이빙의 워크플로우와 보존처리 기술에 관한 연구

## A Study on the Digital Archiving Workflow and Preservation Techniques

이 수 상(Lee Soo-Sang)\*

### < 목 차 >

- |                       |                      |
|-----------------------|----------------------|
| I. 서론                 | 4. 새로운 아카이빙 워크플로우 제안 |
| II. 디지털객체와 아카이빙 워크플로우 | III. 디지털 아카이빙 요소기술   |
| 1. 디지털객체              | 1. 보존처리기술            |
| 2. 디지털객체의 수명주기        | 2. 지원도구              |
| 3. OAIS 참조모델의 보존기능    | IV. 결론               |

### 초 록

디지털보존은 디지털정보를 장기적으로 보유하여 이용자로 하여금 지속적으로 접근이 가능하도록 하는 일련의 행위를 말한다. 이를 위해서는 기술적인 문제를 포함하여 정책적인 문제 기관의 역할 법적 및 저작권 문제, 메타데이터 전략 등 여러 가지 복잡한 세부 과제들을 해결하여야 한다. 본 논문은 이 중에서 디지털보존과 관련된 기술적인 문제를 2가지로 구분하여 정리하고 있다. 먼저, 디지털보존의 기본적인 개념을 형성하는 요소인 디지털객체, 디지털객체의 수명주기, 아카이빙 워크플로우를 시스템의 입장에서 정의하였다. 둘째, 기술보존, 에뮬레이션, 파일재생, 포맷전환, 캡슐화 등 디지털보존에 사용되는 주요한 보존처리 기술을 비교분석하였다.

주제어 : 디지털 보존, 디지털 아카이빙, 디지털객체, 아카이빙 워크플로우, 보존기술전략

### Abstract

The objective of digital preservation is to ensure that digital information survives in usable form for as long as it is wanted. That is to ensure long-term access to digitally stored information. The preservation of digital objects involves a variety of challenges, including policy questions, institutional roles, legal and rights issues, metadata, and technical strategies. In this paper, I introduce the definition of digital object, life-cycle of digital object, and related archiving workflow which are basic concepts of digital archiving. Also I focus on the aspects of current techniques used in digital information preservation including technology preservation, emulation, refreshing, migration, encapsulation.

Key Words : digital archiving, digital preservation, digital object, archiving workflow, preservation techniques.

\* TG인포넷 정보기술연구소 부장(sslee@tginfonet.com)

· 접수일 : 2004. 8. 18 · 최초심사일 : 2004. 9. 1 · 최종심사일 : 2004. 9. 8

## I. 서론

디지털문헌의 아카이빙(digital archiving)은 말 그대로 디지털문헌을 안전하게 보존하는 활동이며, 시간이 경과되어도 그것에 접근할 수 있고 진본을 유지할 수 있도록 하는 모든 행위들을 말한다. 보존문제가 부각하게 된 배경은 정보기술의 발전으로 인하여 디지털 정보의 생산과 유통이 급격하게 증가하였기 때문이다. 디지털문헌은 비트열(또는 바이트열)의 형식으로 CD-ROM, DVD, 디스크 등과 같은 저장매체에 디지털파일로 저장된다. 디지털파일의 비트열을 보존한다는 것은 디지털파일을 보다 안정된 저장매체에 저장하는 일이다. 그러나 만약 저장매체가 손상되거나, 관련된 H/W와 S/W의 기술적인 퇴화가 되기 전에 디지털파일을 다른 저장매체로 전환시켜 놓지 않으면, 디지털정보는 당장이라도 손실이 생기게 된다. 그러므로 10년을 넘기기가 힘들다는 저장매체의 수명이 다하기 전에 미리 디지털정보를 전환하거나 변환하는 작업이 필요하다.

정리하면, 디지털보존(digital preservation)이라고 하면, 디지털문헌의 무결성(integrity)을 상실하지 않고, 저장매체의 손상 및 퇴화(media obsolescence) 또는 H/W 및 S/W의 문제로 인한 기술적 퇴화(technology obsolescence)의 위협으로부터 디지털정보를 안전하게 보호하여 유용한 형태로 존속토록 하는 일련의 관리활동인 셈이다.<sup>1)</sup>

NINCH 워킹그룹(NINCH Working Group on Best Practices)은 2가지 유형에서 장기적 접근 문제를 해결하여야 한다고 주장하고 있다.<sup>2)</sup> 첫째, 기계접근(machine accessibility)의 문제이다. 이것은 디지털문헌의 비트열을 유지하는 것을 말한다. 둘째, 인간접근(human accessibility)의 문제로서, 디지털문헌의 내용을 유지하여 인간의 육안으로 식별이 가능하여야 한다는 것이다. 이것은 디지털보존이 레코드의 비트열만을 단순히 보존한다는 것이 아님을 의미한다. 비트열을 인간의 육안으로 식별할 수 있는 형태로 해석할 수 있어야 한다. 해석할 수 없다면 그것은 0과 1의 의미없는 비트열일 뿐이다.<sup>3)</sup>

이용자 서비스의 관점에서 본 디지털 보존은 이용자에게 디지털문헌의 장기적인 접근을 제공하기 위한 것이다. 이를 위해서는 디지털문헌이 장기적으로 보존되어 있어야 한다. 그러므로 디지털보존은 시스템 내에서 디지털문헌을 손실이나 훼손없이 보유하고, 이용자에게 접근이 되는

1) H/W와 S/W의 기술적 퇴화 문제는 H/W기능 에러 등 물리적 퇴화 새로운 매체의 등장과 같은 기술적 변화 등에 의해 발생하며, 저장매체의 손상 및 퇴화 문제는 H/W와 S/W의 기술적 퇴화, 온도와 습도 같은 저장환경의 영향, 그리고 소유자의 조작실수 등에 의해 발생한다.

2) NINCH Working Group on Best Practices, "The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials," *The National Initiative for a Networked Cultural Heritage* (October 2002).

<<http://www.nyu.edu/its/humanities/ninchguide/index.html>> [cited 2004. 7. 10]

3) Digital Bewaring, "Migration : Context and Current Status," *Digital Preservation Testbed White Paper* (2001), 4쪽 <<http://www.digitaleduurzaamheid.nl/bibliotheek/docs/Migration.pdf>> [cited 2004. 7. 10]

환경을 제공하며, 그리고 문헌의 내용이 제대로 파악(이해)될 수 있도록 보장해야 한다. 현 시점에서 디지털문헌을 관리하고 접근을 제공하는 것보다, 장기적인 보유와 접근을 제공하기 위해서는 여러 가지 다른 행동들이 필요하다.

그러므로 이 논문은 디지털보존의 여러 가지 논의들 중에서 기술적인 문제에 한정하여 그것의 구체적인 특성과 요건을 살펴볼 것이다. 아카이빙의 대상과 접근방법을 정의하고 활동과정과 관련된 업무기능을 도출하며, 기술적인 보존전략을 구현하는 주요한 방식들을 비교검토하는 등 디지털 아카이빙의 요소기술에 대하여 체계적으로 접근하고자 한다.

## II. 디지털객체와 아카이빙 워크플로우

### 1. 디지털객체

디지털 아카이빙에 대한 구체적인 개념과 작업내역을 살펴보기 전에 아카이빙의 대상이 무엇인지 살펴보는 것은 중요하다. 이것은 아카이빙 처리의 기본 단위이며, 모든 관련된 개념들을 규정하는 기준이 되기 때문이다. 결론적으로 말하면, 아카이빙의 대상은 디지털 콘텐츠 그 자체이며, 보존을 위해서는 각종 보존과 관련되는 정보가 필요하다. 디지털 콘텐츠는 디지털문헌을 구성하는 비트열의 디지털파일이며, 디지털 저장매체에 저장된다. 디지털파일을 담고 있는 저장매체(예를 들어 CD, 디스크, 테이프 등)가 있는 경우 해당 매체는 보존대상이 아니다. 특정한 컴퓨터와 관련된 S/W에 의해 읽히고, 저장되고, 유통되는 디지털파일을 디지털객체(digital object)라고 한다.

한편, 디지털 콘텐츠의 보존과 관련된 동적(행위 또는 작업)에 필요한 각종 정보를 보존 메타데이터(preservation metadata)라 한다. 결국 아카이빙의 1차적인 대상은 디지털객체이며 보존 메타데이터도 2차적인 보존대상이 된다. 디지털 아카이빙 시스템에서에서 디지털객체와 보존 메타데이터는 합쳐서 정보객체(information object)라고 부르고 있다.

디지털객체는 디지털 콘텐츠(실물문헌 또는 디지털문헌)이며, 단일의 콘텐츠객체를 가지는 단일객체(single object)와 서로 종속적인 복잡한 콘텐츠객체(복수의 파일 등)가 하나의 객체를 구성하는 복합객체(complex object)로 구분할 수 있다.<sup>4)</sup> 유형을 구분한다면, 도서관객체, 학습객체

4) 객체그룹(object groups)이라는 유형도 있다. 100편의 편지 그룹이 하나의 플로피 디스켓에 저장되어 있는 경우처럼 복합객체이면서 독립적인 파일들의 그룹으로 구성되는 경우이다. 이 객체를 실제적으로 처리하기 위해서는 100개의 단일객체들로 나누어지든지, 각각 25편의 편지로 구성되는 4개의 그룹으로 나누든지, 하나의 논리적 객체를 구성하든지 해야 한다. Sam Searle, Dave Thompson, "Preservation Metadata : Pragmatic First Steps at the National Library of New Zealand," *D-Lib Magazine*, Vol.9, No.4(2003).. <<http://www.dlib.org/dlib/april03/thompson/04thompson.html>> [cited 2004. 7. 20]

박물관객체, 기록물객체 등이 해당된다. 그러므로 보존 메타데이터는 디지털객체를 인간이 이해할 수 있도록 하는 기술적인 정보, 검색 및 접근에 필요한 서지정보, 각종 보존행위를 수행하기 위한 보존정보, 저작권정보 등이 포함된다.

정보모델링 관점에서 보면, 메타데이터는 디지털객체의 속성을 나타내는 정보라고 할 수 있다. 일반적으로 디지털객체는 컴퓨팅 환경에서 디지털문헌을 정의할 때 사용하는 개념으로서 속성과 행위로 설명된다. 이 중에서 속성은 서명, 저자, 출판년 등 색인 또는 목록 정보, 파일의 크기나 매체 유형 등과 같은 기술적인 정보, 또한 보존에 필요한 정보, 저작권관련 정보, 또는 시스템에서의 각종 관리적 목적의 정보와 같은 메타데이터로 표현된다. 행위는 수집, 조직, 저장, 배포 등과 같은 처리동작으로서 시스템의 주요 기능을 말한다. 보존의 관점에서 정리하면, 앞에서 정리한 보존활동에 필요한 모든 작업(또는 동작)이 시스템의 행위인 것이다. 한편, 디지털객체의 인스턴스(instance)는 학술논문, 학위논문, 연구보고서, 단행본 등과 같은 단순한 형태의 디지털문헌(e-prints), 웹문서, 강의 및 학습자료, 박물관자료 등과 같은 멀티미디어 형태의 자료이다.

정리하면, 디지털 보존시스템에서 디지털객체는 보존활동에 필요한 각종 속성정보인 보존 메타데이터와 관련된 보존행위들로 설명된다. 여기서 디지털객체와 보존 메타데이터를 시스템 아키텍처의 입장에서 모델링하는 방법 즉 정보모델링(information modeling)이 필요하다. 디지털 보존시스템과 관련하여 현재 알려진 정보모델링 방법은 OAIS 정보객체 모델<sup>5)</sup>과 METS 스키마<sup>6)</sup>의 2가지가 있다. OAIS 정보객체 모델은 시스템에서 디지털객체와 보존 메타데이터를 구성하는 각종 정보의 관계를 모델링하는 말 그대로의 모델링 프레임워크이다. 그러나 METS 스키마는 모델링 개념이기 보다는 보존시스템에 필요한 디지털객체와 모든 유형의 메타데이터를 연결하여 인코딩하고, 교환하는 구현 프레임워크 또는 메타데이터 패키징 모형이다. 개념적으로 상통하는 측면이 있지만, 사용하는 용어, 역할, 방법 등에 있어 차이가 있다.<sup>7)</sup>

## 2. 디지털객체의 수명주기

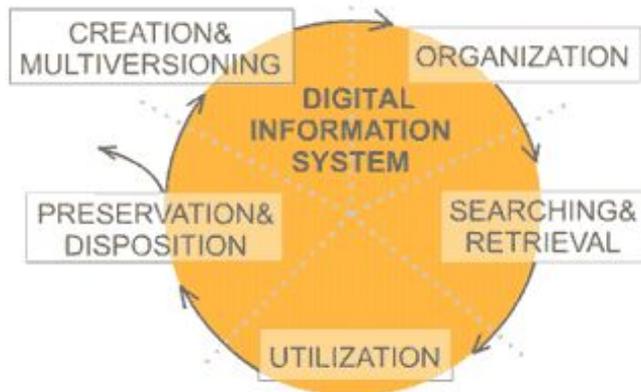
디지털 아카이빙에 필요한 제반 활동은 디지털객체의 수명주기(life-cycle)를 기반으로 파악할 수 있다. 수명주기의 제반 단계에서 보존에 필요한 각종 활동들을 규정할 수 있기 때문이다. 그러므로 수명주기는 디지털객체의 보존 워크플로우(workflow)를 정리하는 기준이 된다. 스위트랜드(Sweetland-Gilliland)는 <그림-1>과 같이 디지털정보시스템에서 디지털객체의 수명주기를 생

5) The OCLC/RLG Working Group on Preservation Metadata, "A Metadata Framework to Support the Preservation of Digital Objects," (2002), 6쪽, <<http://www.oclc.org/research/pmwg>> [cited 2004. 7. 20].

6) <http://www.loc.gov/standards/mets/>

7) 정보모델링 그 자체만으로 하나의 커다란 논제이며, 보존 메타데이터와 함께 논의될 사안이다

성-조직-검색-활용-보존의 단계로 나누어 설명하고 있다<sup>8)</sup>



<그림-1> 디지털객체의 수명주기

생산(creation and multi-versioning) 단계는 디지털객체를 디지털 형태로 생성하거나 디지털포맷으로 변환하여 디지털정보 시스템에 등록하는 단계이다. 동일 객체에 대하여 복수의 버전을 만드는 작업은 보존, 연구, 배포, 또는 제품개발의 목적을 위한 것이다. 이 경우 생산자는 디지털객체를 등록할 때 관리정보와 기술정보 메타데이터를 함께 등록한다. 조직(organization) 단계에서는 디지털객체를 디지털정보 시스템에서 활용하기 위한 용도로 조직작업을 수행한다. 즉, 등록, 목록, 색인 과정을 통하여 부가적인 메타데이터가 생성된다. 검색(search and retrieval) 단계는 이용자가 디지털객체를 검색하고, 콘텐츠에 접근하는 과정을 말한다. 활용(utilization) 단계는 검색된 디지털객체를 활용하고, 재활용하고, 수정하는 행위를 말한다. 마지막으로 보존(preservation and disposition) 단계는 파일재생(refreshing), 포맷전환(migration), 무결성 검증(integrity checking) 등을 통해 지속적인 활용을 보장하는 처리를 수행한다. 더 이상 필요가 없을 경우 폐기되기도 한다.

호지(Gail M. Hodge)는 최적의 아카이빙은 정보의 수명주기에 따라 모든 단계의 정보들을 관리하는 것이라고 하면서, 정보 수명주기를 생산-수집-목록/식별-저장-보존-접근의 6 단계로 나누고 있다.<sup>9)</sup>

- 생산(creation) : 저자, 장치, 또는 도구 등으로부터 정보라는 자원을 생산하는 행동
- 수집(acquisition and collection development) : 생산된 디지털객체를 시스템으로 수집하는

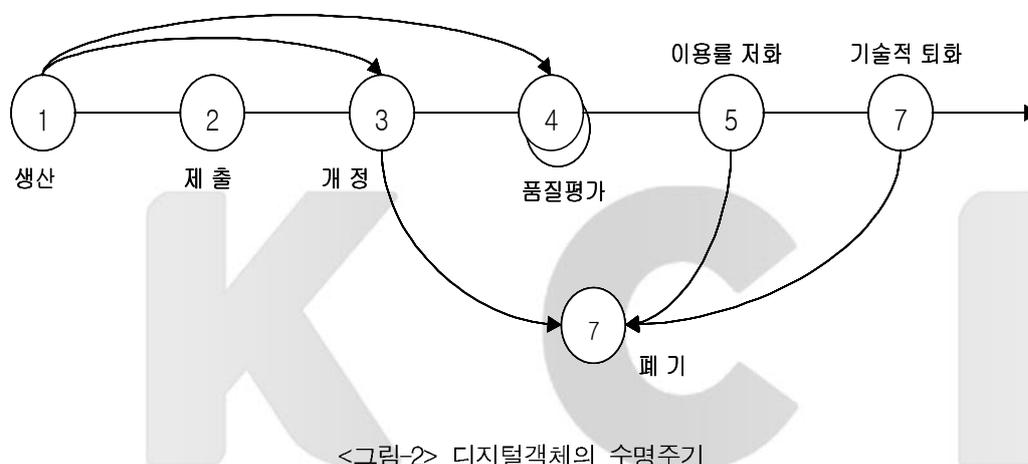
8) Anne J. Gilliland-Swetland, "Setting the Stage," in *Introduction to Metadata : in Pathways to Digital Information* (2000), <<http://www.getty.edu/research/institute/standards/intrometadata/>> [cited 2004. 7. 20]

9) Gail M. Hodge, "Best Practices for Digital Archiving : An Information Life Cycle Approach," *D-Lib Magazine*, Vol.6, No.1(2000), <<http://www.dlib.org/dlib/january00/01hodge.html>> [cited 2004. 7. 20].

단계이며, 규정된 장서정책과 수집절차 등의 기준을 준수

- 목록/식별(identification and cataloging) : 수집된 디지털객체에 고유 식별자를 부여하고 메타데이터로 목록을 생성
- 저장(storage) : 디지털객체를 적절한 저장매체와 파일포맷으로 저장
- 보존(preservation) : 디지털객체의 콘텐츠를 보고 느끼는 그대로의 형태로 보존하는 행위
- 접근(access) : 보존된 디지털객체를 검색하는 단계로서, 적절한 검색방법과 권한관리, 보안 관리 등의 기능이 필요

한편, JISC 보고서에서 디지털문헌의 수명주기를 생산-제출-개정-품질평가-이용률저하-기술퇴화-폐기의 과정으로 세분화하여 설명하고 있다. 이것은 디지털문헌의 전체적인 수명주기에서 발상가능한 활동을 7가지 단계별로 정리한 것으로 <그림-2>와 같이 표현된다.<sup>10)</sup>



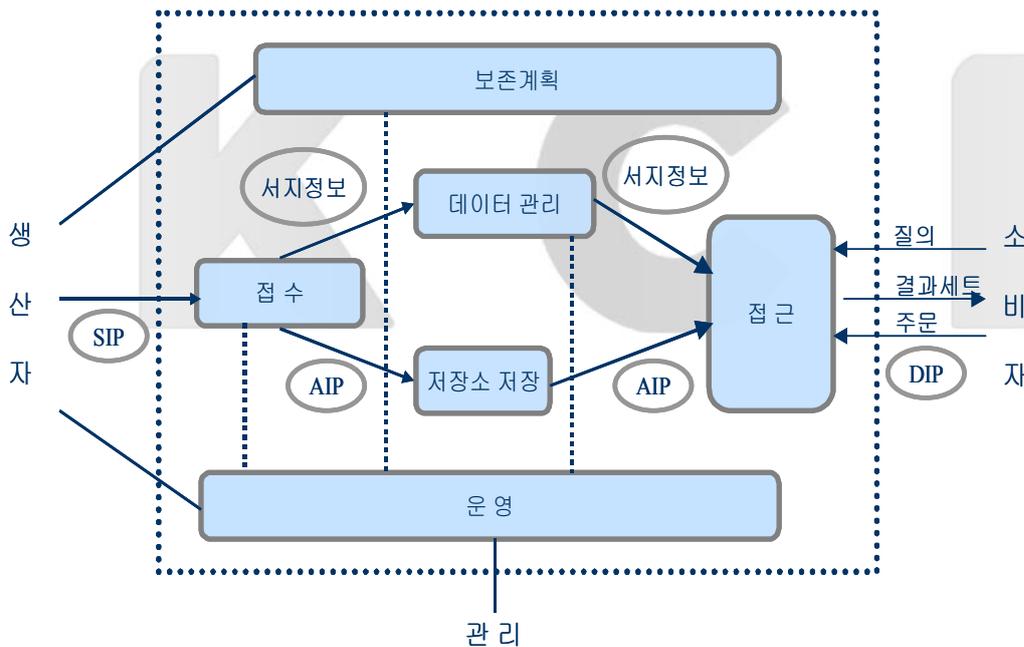
- 생산(creation) : 특정한 S/W 패키지를 사용하여 디지털객체를 생산
- 제출(submission) : 저자 또는 생산자에 의해 디지털객체를 시스템으로 직접제출(self-archiving)
- 개정(revision) : 디지털객체의 버전이 다양할 경우(예: 인쇄전 또는 인쇄후 문헌), 신 버전을 선택하고, 이전 버전의 문헌을 폐기하는 것과 같은 의사결정을 수행
- 품질평가(quality assessment and publication) : 디지털객체의 최종 버전의 수용여부를 판단
- 이용률 저하(end of frequent reading) : 활용빈도가 저하되는 시기

10) James Hamish, et al., "Feasibility and Requirements Study on Preservation of E-prints," *Report Commissions* by the Joint Information Systems Committee (JISC)(2003), 18-20쪽, <[http://www.jisc.ac.uk/uploaded\\_documents/e-prints\\_report\\_final.pdf](http://www.jisc.ac.uk/uploaded_documents/e-prints_report_final.pdf)> [cited 2004. 7. 22]

- 기술적 퇴화(technological obsolescence) : 디지털객체를 생산한 기술적인 환경이 퇴화
- 폐기(withdraw e-print) : 디지털객체를 더 이상 활용하지 않도록 하는 의사결정에 따라 영구삭제

### 3. OAIS 참조모델의 보존기능

OAIS 참조모델<sup>11)</sup>은 디지털 아카이빙 시스템에서 가장 중요하게 언급되는 국제표준이다. 이 표준은 2가지 영역에서 공헌을 하게 된다. 첫째, 그것은 디지털문헌을 설명하는 정보모델링과 둘째, 디지털문헌의 장기적 보존과 검색을 위한 시스템의 표준기능의 요건을 정립한 것이다. 정보 모델링은 디지털문헌의 콘텐츠와 보존활동에 필요한 각종 메타데이터를 패키지로 구성하는 신택스를 제공한다. 또한 이 모델링은 보존 메타데이터의 구성내용과 프레임워크를 제공한다. 디지털 아카이빙 시스템의 표준기능과 요건은 <그림-3>과 같이 디지털문헌의 입수, 저장, 보존, 배포 등에 이르는 수명주기에 따라 구분한 워크플로우와 관련된 보존활동에 필요한 6가지의 표준 기능 요소(functional components)를 말한다.



<그림-3> OAIS 참조모델의 6가지 기능요소

11) CCSDS, "Reference Model for an Open Archival Information System (OAIS)," Red Book CCSDS, 650.0-R-1(2001), <[http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)> [cited 2004. 7. 22].

① 접수기능

생산자가 제출한 정보를 수집하고 저장소에 저장하기 위한 준비 역할을 수행하는 기능이다. 세 부단계를 구분하면, 생산자(제출자)의 정보를 제출용 정보패키지(SIP: Submission Information Package)로 수집하여 OAIS 시스템에 전달하는 기능; 수집된 정보가 손상되지 않았는지, 또는 완전한 것인지 확인하고 점검하는 기능; 제출된 정보를 아카이빙 시스템의 저장 및 관리에 적절한 저장용 정보패키지(AIP: Archival Information Package)로 변환하는 기능; OAIS의 탐색 및 검색 도구에 필요한 서지정보 메타데이터를 추출하고 생산하는 기능; 제출된 정보와 관련 메타데이터인 AIP를 저장소로 전달하는 기능; 서지정보 메타데이터를 데이터관리 기능으로 전달하는 기능 등이다. 요약하면, OAIS 제출기능은 생산자와의 외부 인터페이스이며, 제출된 정보를 수집하여 저장소로 전달하고, 서지정보를 데이터관리 기능으로 전달하는 과정을 말한다.

② 저장기능

이 기능은 제출된 디지털문헌의 비트열과 메타데이터(AIP)를 저장, 보존하는 일을 담당한다. 또한 소비자(이용자)의 요청에 따라 보존하고 있는 AIP를 제공하는 역할도 수행한다. 구체적으로 살펴보면, 접수기능으로부터 AIP 인수, 보존계획에 따른 보존처리 작업을 수행 정기적이고 지속적인 백업작업; 예기치 못한 재난방지를 위한 활동; AIP의 논리적, 물리적 무결성을 보장하는 점검작업; 소비자의 접근기능(저장되어 있는 콘텐츠의 접근요청)에 대응하여 저장된 AIP를 접근기능으로 전달하는 기능 등이 해당된다. 이 기능은 직접적인 외부 인터페이스를 갖고 있지 않으며, 내부적으로 처리된다.

③ 데이터관리

접수기능으로부터 서지정보를 전달받고, 이용자검색에 필요한 제반활동을 수행하는 기능이다. 구체적인 기능으로는 OAIS의 검색도구에 대응하여 저장된 정보를 식별하고 표현하는 서지정보 메타데이터 DB를 유지; 성능데이터 및 접근통계 등과 같은 OAIS 내부시스템 운영에 필요한 관리데이터의 관리 등이 해당된다. 가장 중요한 기능은 데이터베이스를 관리하고 이용자의 질의어를 처리하며, 다른 기능요소들의 요청에 대한 각종 데이터를 생성하고, 새로운 정보가 수집되거나 기존 정보가 수정 또는 삭제될 경우 데이터베이스를 갱신하는 기능이다.

④ 접근기능

이용자가 저장시스템의 콘텐츠를 검색하고 접근하는데 필요한 서비스와 기능을 말한다. 아이템의 위치식별, 검색요청 처리, 아이템의 제공 등 소비자에 대한 서비스를 관리하는 기능이다. 즉 이용자의 요청을 데이터관리 기능에 전달하고, 배포용 정보패키지(DIP: Dissemination Information Package) 형식으로 제공받은 응답을 다시금 이용자에게 전달하며, 아카이빙 시스템의 콘텐츠에 대한 보안

및 접근제어 기능도 수행한다.

⑤ 운영기능

보존업무를 담당한 조직에서 수행하는 각종 운영에 관련된 활동들을 수행한다. 여기에 포함되는 업무는 저장시스템의 S/W를 관리, 모니터링, 개발하는 것뿐만 아니라 제출자(저자)와 제출과 관련된 협약체결(저작권 및 라이선스), 아카이빙 시스템의 정책 및 표준을 수립하는 업무 등이 해당된다. OAIIS의 일상적인 운영상황을 관리하고 다른 서비스 기능들을 조정하는 기능 즉, OAIIS의 내외부 동작의 중심허브 역할을 수행한다.

⑥ 보존계획기능

보존위험을 인식하고, 그것을 처리하는 계획을 개발하는 것과 관련된 하위기능들을 수행한다. (1)관심공동체의 요구사항 점검 : 아카이빙 시스템의 콘텐츠에 관심이 있는 이해당사자들의 그룹이 새로운 기술을 수용하고, 디지털문헌을 보존하는데 영향을 미치는 다른 기술적인 동향들이 있는지 점검하는 역할, (2)기술변화 점검 : 현재 소장하고 있는 디지털문헌이 퇴화되거나 접근에 어려움이 예상되는 기술적 변화를 부단히 인식하는 등의 기술기능을 점검하는 역할, (3)보존전략 및 표준 개발 : 현재와 미래의 요건으로 알려진 보존에 관한 전략 및 표준을 개발 (4)패키징 설계와 포맷전환 계획 개발 : 파일포맷, 메타데이터, 문서화에 관한 표준들을 수용하는 기능 등이다. 이러한 표준들을 제출과정에 적용하는 도구를 개발하고 기술을 정의하는 기능도 여기에 포함된다. 정리하면, 이 기능은 OAIIS의 보존전략과 밀접한 관련이 있으며 기술적 조건에 따르는 외부 환경을 모니터링하며, 각종 기술변화에 따라 정보를 적절히 저장하고, 보존하고 검색하도록 한다.

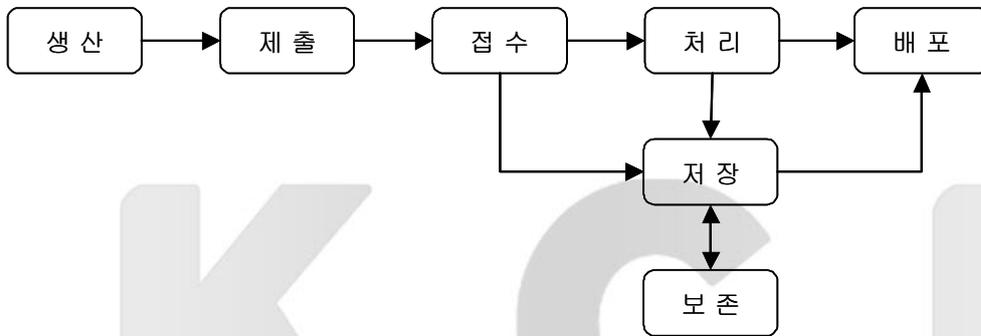
결국 OAIIS 참조모델은 6가지 표준적인 기능요소로 구성되며, 디지털문헌을 장기적으로 보존하고 관련 공동체에서 활용할 수 있도록 해주는 역할을 수행한다. 그러므로 OAIIS 표준모형을 준수하는 디지털 아카이빙 시스템은 이러한 기능들을 구현하여야 한다.

#### 4. 새로운 아카이빙 워크플로우 제안

본 논문에서는 앞서 제시한 디지털객체의 수명주기와 OAIIS 참조모델을 바탕으로 하여 보존시스템의 주요한 업무처리와 기능모듈의 중심이 되는 아카이빙(또는 보존) 워크플로우를 <그림-4>와 같이 생산에서부터 배포까지 전체 7단계로 구분하여 제안한다.<sup>12)</sup> 앞서 살펴본 수명주기 사례

12) 수명주기에 기반한 디지털 아카이빙의 제반 논의는 다음 자료가 좋은 참조가 된다. Gale M. Hodge, "Best Practices for Digital Archiving : An Information Life Cycle Approach," *D-Lib Magazine*, Vol.6, No.1(2000), <<http://www.dlib.org/dlib/january00/01hodge.html>> [cited 2004. 7. 22].

와 OAIS 참조모델은 이론적이거나 개념적인 워크플로우 모델이므로 실제적인 보존시스템을 구현하고자 할 경우 이를 체계적으로 재구성하여야 하기 때문이다. 그러므로 제안되는 워크플로우는 다음과 같은 특성을 가지며, 보존시스템의 실제적인 설계와 구현의 측면에서 유용성을 가지고 있다. 첫째, OAIS 참조모델의 6가지 표준 기능요소의 관련 부분을 워크플로우의 관점에서 재구성하였다. 둘째, 아카이빙 시스템의 외부 작업이지만 생산작업도 중요하므로 생산과정을 추가하였다. 셋째, 제출과정과 접수과정을 분리하였다. 제출은 제출자 또는 납본자의 입장에서 필요한 행위이며, 접수는 시스템의 관리자(사서 아키비스트 등) 입장에서 이루어지는 행위이므로 행위자에 따라 역할을 구분할 필요가 있기 때문이다. 넷째, 처리과정과 저장과정도 구분하였다. 처리과정과 저장과정의 역할을 명확하게 나누기 위해서이다. 넷째, 각종 보존행위를 수행하는 보존과정을 별도로 하였다. 마지막으로, 배포과정은 검색서비스, 내부의 기존시스템과 연동 및 외부 시스템과 연동을 포함하는 각종 디지털객체의 유통에 관한 활동으로 정의하였다.



<그림-4> 아카이빙 워크플로우

① 생산 단계

디지털객체를 생산하는 작업을 말한다. 여기에는 문헌을 디지털 형식으로 직접생산(born digital) 하는 방식과 기존의 아날로그 문헌을 스캐닝하거나 이미지화 작업을 하는 디지털변환(being digital) 작업 모두가 해당된다. 이 단계는 시스템의 외부활동으로 볼 수 있기에, 보존활동과 무관할 수 있다. 그러나 디지털객체가 생산시점에서 어떠한 기술을 사용하였는가 또는 어떤 S/W 를 이용하였는가에 따라 향후 보존활동에 커다란 영향을 미치게 된다. 생산자(creator)가 파일포맷, S/W와 H/W 등 많은 결정을 하기 때문이다 생산단계에서부터 보존문제를 고려할 수밖에 없으므로 중요한 단계인 셈이다.

② 제출 단계

생산자가 디지털객체를 시스템으로 직접 제출(self-archiving)하거나 관리자의 지원으로 제출하

는(또는 일괄제출) 기능이다. 제출자는 제출자인증, 제출대상 컬렉션(collection) 선택, 제출용 메타데이터 입력, 원문업로드, 제출확인 및 로그아웃 등과 같은 미리 규정된 제출 워크플로우(submission workflow)에 따라 제출작업을 수행하게 된다. 관리자나 다른 대리인이 제출작업을 대신할 수 있다. 어떠한 경우이든 제출되는 정보는 OAIS의 제출용 정보패키지(SIP) 형식으로 구성된다. 즉, 제출용 정보패키지는 제출자가 작성하는 메타데이터(서지정보, 저작권정보 등)와 디지털파일, 그리고 제출자정보, 제출대상 컬렉션 등이 포함된다.

이 과정에서의 가장 중요한 쟁점사항은 제출포맷 즉 디지털파일의 포맷을 제한할 것인가 아니면 모든 유형의 디지털파일을 허용할 것인가 하는 문제이다. 또한 제한할 경우 단일포맷으로 할 것인가 몇 가지 유형의 포맷으로 제한할 것인가 하는 문제도 포함된다. 원칙적으로 제출포맷을 제한해서는 안된다. 디지털원본을 있는 그대로 수집해서 보존해야 하기 때문이다. 그러나 현실적으로는 제출포맷의 제한이 불가피할 수 있다. 제출원본 파일은 제한이 없더라도 서비스를 위하여 개방포맷의 복사파일(예를 들어 PDF 파일)도 함께 제출할 것인지도 결정하여야 한다.<sup>13)</sup> 두 번째 쟁점사항은 단일 개수의 디지털파일만 허용할 것인가 복수의 디지털파일도 허용할 것인가 하는 문제이다. 특히 복수의 파일을 허용할 경우 파일간의 구조정보는 어떻게 표현할 것인가 하는 문제도 중요하다.

정리하면, 제출 워크플로우, 제출되는 디지털파일의 포맷, 제출 메타데이터의 범위와 데이터 내용에 대한 결정이 필요하다. 제출 단계에서 필요한 요소기술은 제출과정의 제반 쟁점사항을 수용하여 설계한 ‘온라인 제출기’(online submission tool)이다

### ③ 접수 단계

제출된 정보객체를 시스템의 입장에서 접수하는 단계이다. 제출된 파일이 정해진 포맷 및 버전인지, 컬렉션의 장서관리 정책(collection management policy)에 부합하는 내용인지 등과 같은 품질평가(quality review)와 제출된 메타데이터의 품질검증(metadata review)의 작업이 필요하다. 특별한 문제가 없는 경우로 확인되면, 이 단계에서는 제출정보(OAIS의 SIP)를 수정보완하는 기능, 제출된 디지털파일로부터 적절한 양의 기술정보를 추출하는 기능 그리고 이러한 정보들을 사용하여 OAIS의 저장용 정보패키지(AIP)를 생성하는 기능 등이 필요하다. 저장용 정보패키지는 저장단계에 저장을 요청하며, 이 중에서 서지정보 메타데이터는 처리단계로 이관하는 등의 데이터 이관작업도 수행한다.

이 단계의 쟁점사항은 정책적인 쟁점사항과 시스템적인 쟁점사항의 두 가지로 구분할 수 있다.

13) 전용파일 포맷을 허용할 경우 디지털콘텐츠를 보존하는데 있어 위험이 생길 가능성이 높지만, 개방표준 포맷으로 변환함으로써 위험도를 최소화할 있다. 그러므로 개방표준 포맷을 채택하는 것은 장기적인 관점에서 원문의 내용에 접근하지 못하는 위험을 최소화할 수 있기에 현재로서는 적절한 방안이라 할 수 있다.

정책적인 쟁점사항은 장서관리 정책에 관한 것으로서, 메타데이터 및 디지털파일의 품질평가자(제출자, 검증자, 승인자, 편집자 등), 평가기준 등에 관한 문제이다. 시스템의 장서관리 정책과 부합하는 수집대상인지를 선정기준(selection level)<sup>14)</sup>에 따라 구분하고, 정해진 포맷이나 버전인지, 내용에 대한 품질평가, 메타데이터에 대한 검증작업 등을 수행하게 된다.

또한 제출된 정보로부터 어떠한 내용과 방법으로 기술정보 메타데이터를 추출할 것인가 하는 것도 중요하다. 만일 시스템에서 디지털파일의 각종 기술적인 재현정보(representation information)<sup>15)</sup>를 디지털포맷 등록기(Digital Format Registry)에서 관리하는 경우, 등록기의 관리방법과 기술정보 메타데이터와 어떻게 관계를 설정할 할 것인가 하는 것도 중요한 문제이다. 등록기의 포맷정보가 바로 기술정보 메타데이터이다.

시스템 측면에서 중요한 쟁점사항은 접수단계에서 확보된 각종 메타데이터를 저장용 정보객체로 구성하는 방식을 어떻게 할 것인가 하는 점이다. 즉, 저장용 정보객체가 METS 정보객체 또는 OAIS 정보객체로 구성하는가에 따라 메타데이터 편집 및 처리 방식이 달라지기 때문이다. 또한 저장용 정보객체에 대한 식별체계를 어떻게 할 것인가 하는 것도 이 단계에서 결정해야 하는 사항이다. 그러므로 이 단계의 요소기술은 포맷정보 추출기, 메타데이터 편집기 그리고 저장용 정보객체를 생성하는 보존 메타데이터 편집기(메타데이터 프레임워크) 등이 있다.

#### ④ 처리 단계

이 단계는 이용자의 정보검색에 필요한 각종 데이터처리 작업을 수행한다. 서지정보 메타데이터 처리(메타데이터 편집기), 검색용 색인데이터를 추출하고 관리하는 기능, 검색 및 화면출력 등과 같은 이용자 서비스요청에 대한 처리, 제출자 관리, 인증 및 접근제어 처리 기능, 그리고 각종 통계처리 기능 등이 해당된다. 이 단계에서는 보존활동과 관련해서는 특별한 쟁점사항은 없으며, 관련된 요소기술도 해당이 없다고 하겠다.

#### ⑤ 저장 단계

저장단계는 접수단계에서 요청한 저장용 정보객체(디지털객체와 메타데이터)를 저장장치에 저장하는 기능을 수행한다. 또한 이용자의 검색 요청에 따라 저장된 정보객체에서 필요한 정보를

14) 현재 디지털 장서의 선정기준은 버클리대학의 디지털도서관에서 제시한 다음의 선정기준을 기초로 논의되고 있다. 선정기준의 유형은 1) Archived(보존대상), 2) Serviced(보존대상은 아니지만 단위기관에서 소장 가능), 3) Mirrored(보존대상은 아니며, 여러 기관에서 소장 가능), 4) Linked(외부에서 소장하고 있는 자원이며, 연계기능으로 통해 서비스 제공).

Berkeley Digital Library Sunsite, <<http://sunsite.berkeley.edu/Admin/collection.html>>

15) 재현정보는 OAIS 모델에서 사용하는 개념으로, 일반적으로는 기술정보 메타데이터(technical metadata)라 한다. 디지털매체로 인코딩된 지적 내용(콘텐츠)에 접근하는데 필요한 지식을 제공하는 것으로서, 보존 메타데이터에서 가장 핵심적인 정보이다.

OAIS의 배포용 정보패키지(DIP) 형식으로 생성하는 역할을 수행한다.

주요 쟁점사항으로는 첫째, 디지털객체의 콘텐츠 파일에 대한 저장방식이다. 콘텐츠를 비트열(또는 바이트열) 형식으로 저장하는 방식과 전용 형식의 포맷으로 저장하는 방식으로 구분이 가능하다. 전자는 이진코드(binary code)로 저장하기에 안전하며 편리한 방식이다. 후자는 XML(또는 SGML)과 같이 디지털 콘텐츠의 텍스트를 있는 그대로 저장하거나, PDF 포맷과 같이 많이 사용하는 포맷으로 저장하는 것이다. 이상적으로는 저장포맷이 개방표준형인 XML 포맷이지만 현실적으로는 PDF포맷과 같은 전용포맷을 사용할 수밖에 없는 경우도 있다. 둘째, 디지털 콘텐츠와 저장용 메타데이터를 분리하여 저장하거나 저장용 정보객체에 디지털 콘텐츠까지 포함하여 저장할 수도 있다. 셋째, 파일의 인증정보(체크섬 등), 압축저장, 암호화 적용, 백업저장 등과 같은 처리도 이 단계에서 요구되고 한다.

#### ⑥ 보존 단계

이 단계에서는 디지털파일의 보존에 필요한 여러 가지 작업을 수행한다. 보존작업은 시스템에 따라 다르지만, 가장 많이 논의되는 방식은 파일전환(migration)과 에뮬레이션(emulation) 방식이다. 이와 같은 보존작업을 동작하기 위한 스케줄러, 또는 보존작업을 유발하는 기술감시기능(각종 기술적인 변화를 감시하는 기능)과 같은 작업도 필요로 한다. 또한 보존작업에 따라 원문파일의 변경정보를 기록유지하고, 절차(보존연한)에 따라 필요없는 파일을 폐기할 수도 있다. 이 단계에서는 보존작업에 필요한 지식데이터베이스로서 재현정보시스템이 필요하다. 디지털 파일 포맷에 대한 여러 가지 의미있는 정보를 유지하고 있어 파일전환이나 에뮬레이션이 동작할 때 중요한 지식을 제공하게 된다.

그러므로 주요 쟁점사항으로는 어떠한 보존작업 방식을 지원할 것인가 하는 것이다. 이것은 보존 시스템이 전략적인 차원에서 결정해야 하는 사항이므로 보존전략(preservation strategies)이라고 부르고 있다. 실질적인 보존작업을 실행하는 방식(이용자 요청, 기술감시기 요청, 주기적인 방식 등)에 대한 선택도 중요하다. 이 단계의 요소기술은 보존처리기, 기술감시기, 재현정보시스템 등이 있다.

#### ⑦ 배포 단계

이 단계는 저장되어 있거나 보존작업에 의해 유지되고 있는 디지털 콘텐츠를 이용자의 요청에 따라 제공해 주는 역할을 수행한다. 보존활동과는 무관할 수 있지만, 보존전략(즉, 이용자 요청에 의한 포맷전환, 또는 가상의 통합뷰어 사용 등)에 따라 관련이 있다. 배포는 이용자 검색과 정보객체 배포의 2단계로 구분할 수 있다. 이용자의 검색은 일반검색, 고급검색, 브라우징, 통합검색 등 다양한 방식으로 가능하다. 검색된 정보의 디지털 콘텐츠는 배포용 정보객체로 구성되어 이용자에게 배포(제공)해 준다. 즉, OAIS의 배포용 정보패키지인 DIP를 만들어 제공한다.

검색된 디지털 콘텐츠에 접근하기 위해서는 콘텐츠의 식별자에 대한 해석 작업 (identifier handling)과 다운로드 작업, 필요시 반출 작업 등이 필요하다. 디지털객체를 외부와 유통하는 기능도 이 단계에 포함된다. 기존시스템과 연동, 외부 시스템에 정보를 공개하기 위한 OAI 프로토콜 수용, 정보객체의 교환 등과 같은 작업이 해당된다.

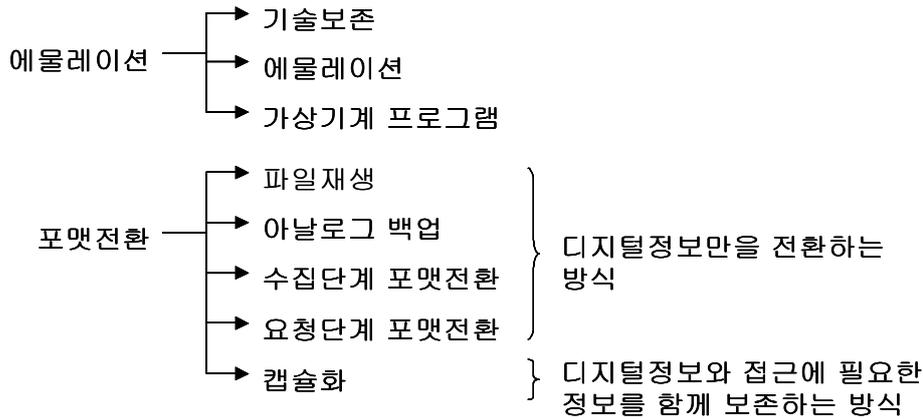
주요 쟁점사항으로는 이용자에게 디지털객체의 지적 내용에 어떻게 접근하도록 하는가 하는 문제이다. 이 문제는 저장/보존된 디지털객체의 내용변환(rendering) 방식을 말한다. 즉, 저장된 원문파일 그대로 제공하거나 또는 포맷전환하여 제공, 공통의 뷰어를 통한 제공 등의 방식이 가능하다. 이 문제는 보존작업 처리와 연관되어 있다. 추가적인 쟁점사항으로는 저작권 제어방식, 표준화된 유통방식(OAI 기반의 메타데이터 및 원문 유통), 일괄 반출 기능 등에 대한 고려도 필요하다. 이러한 작업을 수행하는 기능들 모두가 이 단계의 요소기술에 해당된다

### III. 디지털 아카이빙 요소기술

3장의 보존활동의 워크플로우 단계에서 언급한 요소기술 중에서 실질적인 보존활동과 관련이 있는 요소기술은 보존단계에서 수행하는 각종 보존처리기술과 지원도구이다. 여기서는 에뮬레이션과 포맷전환으로 대표되는 보존처리기술과 이를 지원하는 포맷정보 추출기, 재현정보시스템 등과 같은 각종 도구의 현황과 내용에 대하여 구체적으로 살펴보고 실제적인 보존시스템을 설계하고 구현하려고 할 경우 기술적인 의사결정에 도움을 주고자 한다.

#### 1. 보존처리기술

보존처리기술은 보존활동에서 가장 중요한 요소기술로서, 계획된 보존전략을 수행하는 일련의 기술적 접근방식을 말한다. 아카이빙 시스템 내에서 하나의 보존처리기술만을 사용하지는 않는다. 포맷에 따라, 기술적 환경에 따라 여러 기술을 혼용할 수 있다. 아무튼 현재 가장 많이 알려진 보존전략은 크게 에뮬레이션(emulation) 방식과 포맷전환(migration) 방식의 처리기술로 구분된다. 에뮬레이션은 미래의 사용을 위하여 H/W와 S/W 환경을 구축하는 방식이다. 기술보존 기술 에뮬레이션, 가상기계 프로그램의 3가지 방식이 많이 알려져 있다. 포맷전환 기술은 디지털포맷의 기술적 퇴화에 따른 데이터 손실을 극복하기 위한 방식이며, 이는 다시금 디지털정보만을 전환하는 방식(가능한 개방된 표준포맷으로 전환)과 디지털객체와 그것의 접근에 필요한 정보를 함께 보존하는 방식으로 구분이 가능하다. <그림-5>는 이러한 방식을 요약정리한 것이다.



<그림-5> 보존처리기술 방식

① 기술보존 (technology preservation)

원본을 접근하는데 요구되는 모든 기술(S/W, H/W, O/S, 그리고 이들을 구동하는데 필요한 기술 등)을 보존하는 방식이다. '컴퓨터 박물관(computer museum)'이라 부르기도 하는 이 방식은 우선 비용이 많이 들고, 기술적으로도 쉬운 문제가 아니다.

② 에뮬레이션(emulation)

이 방식은 디지털원본에 적용된 기술적인 조건들에 변경이 있어도 인코딩되어 있는 콘텐츠를 재생할 수 있는 환경을 프로그램으로 만들어 내어 디지털문헌의 접근성을 보장해주는 방식이다. IBM, CAMLEON 프로젝트에서 검토한 방식으로, 아직은 여러 가지 해결하여야 하는 과제들이 많다. 이전의 H/W나 S/W를 유사하게 흉내를 내어 처리하므로, 디지털문헌의 포맷이나 인코딩 방식을 변화하지 않고 원본을 그대로 보존할 수 있는 장점이 있다. 그러나 목표로 하는 컴퓨팅 플랫폼에서 에뮬레이션이 원본재현의 성공률이 높도록 하기 위한 막대한 양의 정보를 유지해야 하며, 수준높은 기술이 적용되어야 한다. 또한 많은 파일포맷들을 처리하더라도 비용측면에서 효율적이어야 한다. 그러므로 에뮬레이션을 위한 메타데이터를 충분히 유지하여 에뮬레이션 S/W와 시스템의 활용에 도움이 되어야 한다.

③ 가상기계 프로그램(virtual machine software)

에뮬레이션의 한 변형으로서, 디지털파일의 내용해석을 미래의 범용 가상컴퓨터(UVC: Universal Virtual Computer)의 기계언어로 처리할 수 있도록 프로그램을 작성하여 해결하려는 시도이다. 범용 가상컴퓨터는 IBM에서 설계한 플랫폼이다 이론적으로는 디지털보존의 장기적 쟁점들에 부합하는 유일한 전략으로 인식할 수 있지만, 구현이 쉽지 않다는 문제가 있다.

④ 파일재생(refreshing)

디지털객체를 동일한 또는 유사한 포맷으로 복사하는 방법이다. 통상적인 백업작업과 유사하므로 특별히 장기적 보존을 위한 기술이라고 말하기 곤란한 측면이 없지는 않다. 그러나 H/W의 장애나 저장매체의 문제로 인하여 디지털객체의 물리적 내용을 변화없이 그대로 복사하는 비트열 복사(bit-stream copying)와 하나의 저장매체에서 다른 매체로 디지털정보를 복사하는 작업(4mm DAT 테이프에서 QIC 테이프로 파일을 복사)이 여기에 해당된다.<sup>16)</sup>

⑤ 아날로그 백업(analog backup)

디지털 객체를 보존성이 높은 매체인 아날로그 형태로 변환하는 방식이다. 인쇄형태로 출력하든지, 디지털 이미지 파일을 마이크로 형태로 변환하든지 하는 경우이다. 이 방식은 다소 고전적인 방식으로 디지털객체의 콘텐츠를 보존하고 기술적인 퇴화에 대응할 수 있다. 그러나 디지털문헌에 대한 기능적 또는 관련 행동에 관한 정보의 손실이 발생할 수 있다.

⑥ 점수단계 포맷전환(migration on ingest)

디지털객체를 보존시스템으로 접수하는 시점에서 전용포맷의 파일을 공개된 표준포맷으로 전환할 때 사용하는 방식이다. 현재로서는 가장 쉽게 사용하는 방식이다. 이 방식이 성공하려면 현재의 표준포맷이 오랫동안 존속하여야 한다. 그러나 표준포맷이 생존하지 못하거나, 사람들의 선호가 바뀔 때마다, 지속적으로 포맷을 전환해 주어야 하는 문제가 있다. 아무튼 이 방식은 가장 실용적이라고 평가받고 있다. 그러나 표준포맷에 대한 합의가 쉽지 않고, 표준포맷이 변경되거나 파일이 퇴화될 경우, 예상치 못하는 변환처리가 필요하며 이 경우 커다란 비용이 소요될 가능성이 높다.

⑦ 요청단계 포맷전환(migration on request)

접수단계에서 포맷전환할 경우 나타나는 비용과 장기적인 존속성 문제의 해결책으로 제시된 방식이다. CAMLEON 프로젝트에서 사용한 방식이며, 보존대상의 디지털객체를 이용자가 접근하여 활용하는 시점에서 유용한 포맷으로 전환하게 된다. 그러므로 디지털객체는 필요한 관련 도구와 함께 오랫동안 변경되지 않고 보존하고 있어야 한다. 유사한 파일포맷으로 된 디지털객체들이 보존되었을 경우, 포맷전환 도구가 쉽게 처리할 수 있어 다소 경제적인 방식이다. 포맷전환 도구들을 오랫동안 보존할 수만 있다면, 원본과 이용자에게 보이는 콘텐츠 사이에 단지 하나의

16) 포맷전환과 파일재생은 파일의 디지털정보를 주기적으로 전환하거나 복사하는 방법이다. 그러나 그 차이를 구분해 보면, 전자는 한 H/W 및 S/W의 사양에서 다른 사양으로 또는 컴퓨터 기술의 한 세대에서도 다음 세대로 디지털파일을 변환하는 방식이다. 후자는 유사하거나 새로운 저장매체로 디지털파일의 비트열을 주기적으로 복사하여 재생하는 방법이다. 즉, 포맷전환은 제반 기술적인 환경의 이유로 파일을 전환하는 경우이고, 파일재생은 저장매체의 문제에 대비한 파일전환인 것이다

단계만 있으면 되므로, 다른 포맷전환 방식보다 정확성이 높을 수 있다.

이 방식의 성공요인은 초기에 구현할 경우 품질이 얼마나 좋은가에 달려있다. 즉, 미래의 요청에 처리를 잘 할 수 있도록 초기에 잘 구현하여야 한다는 것이다. 만일 초기 구현에 문제가 있으면, 시간이 경과할수록 정확하고 효율적인 보존이 어려워진다. 이 전략은 중장기에 전체적인 보존비용을 경감시킬 수 있는 장점이 있다. 그러나 초기 개발단계에서 상대적으로 많은 구현노력이 요구되기도 한다.

⑧ 캡슐화(encapsulation)

포맷전환 방식과 달리, 캡슐화 방식은 디지털문헌의 원본을 그대로 유지하며 원본을 해석하는데 필요한 세부절차에 관한 정보를 함께 보존한다. 이 생각이 확장되어 실행가능한 프로그램을 구성한 것이 바로 가상기계 프로그램이다.

이경호(Lee Kyong-ho)는 대표적인 보존기술의 3가지 방식(에물레이션, 포맷전환, 캡슐화)의 장단점과 디지털정보의 유형과 복잡성, 포맷, 사용 방식에 따라 적합한 보존기술을 선택하는 의사결정적인 다이어그램을 제안하고 있다. 이를 소개하면 <표-1>과 같다.<sup>17)</sup>

<표-1> 3가지 보존기술의 장단점

구분	장점	단점	대상 영역
에물레이션	-보고 느끼는 형태 그대로 유지	-에물레이션 기술사양의 구축이 복잡함 -많은 양의 정보를 보존하여야 함 -오래된 소프트웨어 사용	-응용 소프트웨어 -실행파일을 포함하는 복잡한 디지털자원 -충분한 정보가 부족한 자원 -미래의 가치와 활용이 불확실한 자원
포맷전환	-원본의 응용 환경을 보유할 필요가 없음 -적극적인 접근과 관리 활동을 지원	-많은 비용 소요 -정보내용의 퇴화 -보존 메타데이터의 부족 -관리자에 의한 지속적인 관심 요구	-과학데이터와 DB처럼 적극적인 접근과 관리가 필요한 자원 -충분히 알려진 포맷의 자원
캡슐화	-보존정보를 유지	-포맷에 관한 지식을 보존하여야 함 -디지털정보를 추출하는 기능 필요	-적극적인 접근과 관리가 요구되지 않는 자원 -잘 알려지지 않은 포맷의 자원

17) Lee Kyong-ho et. al., "The State of the Art and Practice in Digital Preservation," *Journal of Research of the National Institute of Standards and Technology*, Vol.107, No.1(2002), 93-106쪽, <<http://nvl.nist.gov/pub/nistpubs/jres/107/1/j71lee.pdf>>

## 2. 지원도구

### ① 포맷정보 추출기

포맷정보 추출기는 디지털객체의 접수단계에서 재현정보를 포착하는 기능을 수행한다. 파일포맷을 식별하고 인증하며, 관련되는 기술정보 메타데이터를 자동으로 추출한다. 접수단계에서 디지털보존과 관련된 기술적인 정보를 추출함으로써, 향후 처리될 보존작업에 유용한 정보로서 활용된다. 각각의 기능은 모듈식으로 구현되어 필요시 플러그인 되도록 해야 한다

### ② 기술감시기

이 기술은 재현정보시스템의 주요한 내부기능으로 구현되거나, 외부에서 구현된 기능을 활용하여 구현된다. 재현정보와 관련된 변환기능의 역량을 모니터링, 기술퇴화로 인하여 재현정보가 현재의 정보를 유지할 수 없을 때 그 사실을 공지(alert)하게 된다. 따라서 재현시스템은 이와 같은 적절한 기술감시 기능을 갖추어 재현정보가 현재의 상태를 유지하도록 보장할 수 있어야 한다. 기술감시 기능은 자동으로 점검하거나 수동으로 점검할 수 있다.

### ③ 재현정보시스템

재현정보시스템은 디지털객체에서 재현정보를 추출하여 보완하여 저장하고 이용하는 메커니즘이다. 장기적 보존에 중요한 역할을 수행하는 기능이지만 아직은 관심이 적은 편이다. 현재 논의되고 있는 방식으로는 2가지가 있다. 디지털포맷등록기(digital format registry) 방식<sup>18)</sup>과 재현네트워크(representation network) 방식이 그것이다. 최근에는 METS(Metadata Encoding and Transmission Standard) 스키마에서 재현정보를 기술정보 메타데이터로 관리하는 방식도 제안되고 있다.

- 디지털포맷등록기 : 파일포맷에 관한 간단한 데이터베이스를 유지하는 방식이다. 파일포맷을 유형별로 나누어 그것에 대한 상세한 재현정보를 관리하고 있다. 이 경우 2가지 유형으로 구분이 가능하다. 단위기관 차원에서 하나의 데이터베이스를 구축하여 관리하는 경우와 인터넷을 통하여 원거리 접근을 허용하는 네트워크형 방식이 그것이다. 가장 많이 알려진 적용사례로는 파일포맷의 고유식별자와 해당 포맷의 재현정보를 연결한 등록기로서 GDFR(Global Digital Format Registry)이 있다.
- 재현네트워크 : OAI에서 제안한 방식이며, 기술정보 메타데이터의 네트워크라 할 수 있다. 특히 복수의 파일로 구성된 디지털객체를 처리하는데 효율적인 방식으로 알려져 있다. 특정한 파일포맷에 관한 많은 정보는 다른 유형의 포맷과도 공유되는 정보가 많으므로, 포맷, 인

18) 이 방식은 파일포맷등록기(file format registry)라고 부르기도 한다.

코딩방식, 변환방식을 설명하는 문서 등을 네트워크 노드로 구성하여 상호간 연결하는 방식이다. 현재 CEDARS에서 적용되고 있는 사례이다.

- METS의 기술정보 메타데이터 : 보다 상세하게 구조적인 재현정보를 표현하며, 현재 여러 기관 및 공동체에서 관심을 보이고 있으며, 관련 표준 메타데이터가 개발되고 있다.<sup>19)</sup>

#### ④ 내용변환기

비트스트림을 의미있는 정보로 재현하고, 순수 데이터에 있는 지적내용에 접근하도록 하는 기능이다. 내용변환은 디지털문헌을 활용하려는 이용자의 관점에서 본 변환기술이므로, 이를 구현하는 실제적인 방식은 보존처리기에서 언급한 매체전환 방식과 에플레이션의 2가지 방식으로 크게 구분한다.

## IV. 결론

본 논문에서는 디지털보존의 여러 가지 논의들 중에서 기술적인 문제에 한정하여 그것의 구체적인 특성과 요건을 살펴보았다. 디지털 아카이빙의 기본적인 처리단위인 디지털객체와 그것의 수명주기, 아카이빙을 위한 워크플로우에 대한 참조모델과 새로운 제안모델을 통하여 아카이빙의 대상과 관련 업무 영역과 고려사항을 정리하였다. 그 다음에는 디지털 아카이빙의 기술적인 보존 전략과 그것을 구현하는 여러 가지 방식의 특징을 비교검토 하였다.

디지털 정보환경에서 디지털 아카이빙의 문제는 아무리 강조하여도 지나치지 않는 사안이다. 교육, 연구, 문화 등의 영역에서 디지털자원의 생산은 증가하고 있지만 불행하게도 전체가 통합되고 접근이 용이한 디지털 아카이빙 체계가 제대로 마련되어 있지 못한 것이 현실이다. 이러한 상황에서 본 논문은 디지털 아카이빙에 관련된 기술적인 과제와 관련된 기본적인 조건만을 대상으로 한 한계가 있다.

그러나, 디지털 아카이빙 또는 보존의 문제는 한 기관에서 모든 것을 감당할만한 수준의 과제가 아니다. 국가적 수준에서 해결하든지, 영역별로 협력체를 구성하여 접근하든지 하는 거시적인 접근이 필요하다. 또한 기술적인 문제, 정책적인 문제, 제도 및 법적 문제, 비용 및 운영모델 문제 등과 같이 미시적인 영역에서 검토하고, 해결하고, 결정해야 하는 과제들도 다양하다.

그러므로 한 가지 제언을 한다면, 도서관 및 박물관, 문화기관, 교육기관, 정부기관 등의 관련 당사자들(stakeholder community)을 중심으로 국가적 차원에서 ‘디지털 아카이빙 연구센터와 같

19) <http://www.loc.gov/standards/mets/>

은 조직을 꾸려서 협력적으로 대응하자는 것이다. 이 센터에서는 외국의 관련사례를 조사연구하고, 국내의 아카이빙 기반환경에 대한 세밀한 조사작업을 수행하며, 정책과 제도적인 문제 보존 메타데이터의 정의, 요소기술을 적용한 다양한 테스트베드(testbed) 프로젝트를 실행하는 등을 일을 수행하여야 한다. 이러한 작업은 디지털 아카이빙에 관한 다양한 경험과 기술, 지식을 축적하는 기반이 되며, 국내의 환경에 적합한 아카이빙 모델과 체제를 정립하고, 각종 표준 및 인증을 위한 활동을 수행하며, 관련 전문가 집단을 양성하고 교육하는 기반이 될 것이다.

**<참고문헌은 각주로 대신함>**

